



languages

Special Issue Reprint

Grammaticalization across Languages, Levels and Frameworks

Edited by
Renata Engfels, Marlies Jansegers and Tom Bossuyt

mdpi.com/journal/languages



Grammaticalization across Languages, Levels and Frameworks

Grammaticalization across Languages, Levels and Frameworks

Editors

Renata Enghels

Marlies Jansegers

Tom Bossuyt



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editors

Renata Enghels
Ghent University
Ghent
Belgium

Marlies Jansegers
Ghent University
Ghent
Belgium

Tom Bossuyt
Ghent University
Ghent
Belgium

Editorial Office

MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Languages* (ISSN 2226-471X) (available at: https://www.mdpi.com/journal/languages/special_issues/TG550Y4604).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-2101-3 (Hbk)

ISBN 978-3-7258-2102-0 (PDF)

doi.org/10.3390/books978-3-7258-2102-0

© 2024 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license.

Contents

| | |
|---|------------|
| About the Editors | vii |
| Preface | ix |
| Renata Enghels, Marlies Jansegers and Tom Bossuyt Introduction to the Special Issue Grammaticalization across Languages, Levels, and Frameworks Reprinted from: <i>Languages</i> 2024, 9, 254, doi:10.3390/languages9070254 | 1 |
| Catalina Fuentes Rodríguez From Peripheral Structure to Discourse Operator: <i>No Veas</i> Reprinted from: <i>Languages</i> 2023, 8, 254, doi:10.3390/languages8040254 | 8 |
| Sandra Paoli Gradualness of Grammaticalization and Abrupt Change Reconciled: Evidence from Microvariation in Romance Reprinted from: <i>Languages</i> 2024, 9, 138, doi:10.3390/languages9040138 | 23 |
| Felicitas Otte, Anke Müller, Sabrina Wähl and Gabriele Langer Numeral Incorporation as Grammaticalization? A Corpus Study on German Sign Language (DGS) Reprinted from: <i>Languages</i> 2023, 8, 153, doi:10.3390/languages8020153 | 39 |
| Mar Garachana and María Sol Sansiñena Combinatorial Productivity of Spanish Verbal Periphrases as an Indicator of Their Degree of Grammaticalization Reprinted from: <i>Languages</i> 2023, 8, 187, doi:10.3390/languages8030187 | 71 |
| Julio Torres Soler and Renata Enghels From Motion to Causation: The Diachrony of the Spanish Causative Constructions with <i>traer</i> (‘Bring’) and <i>llevar</i> (‘Take’) Reprinted from: <i>Languages</i> 2023, 8, 122, doi:10.3390/languages8020122 | 96 |
| Mihaela Iliaoaia Constructing Meaning: Historical Changes in MIHI EST and HABEO Constructions in Romanian Reprinted from: <i>Languages</i> 2024, 9, 38, doi:10.3390/languages9020038 | 120 |
| Chris A. Smith Productivity from a Metapragmatic Perspective: Measuring the Diachronic Coverage of the Low Level Lexico-Grammatical Construction Have the N (Body Part/Attitude) to ↔<Metapragmatic Comment> Using the COHA Reprinted from: <i>Languages</i> 2023, 8, 92, doi:10.3390/languages8020092 | 143 |
| Aina Torres-Latorre and Andreu Sentí Clitic Placement and the Grammaticalization of the Future and the Conditional in Old Catalan Reprinted from: <i>Languages</i> 2023, 8, 182, doi:10.3390/languages8030182 | 166 |
| Eleonora Morei Not Only Anteriority in the Past: The Functions of the Pluperfect in Spoken Italian Reprinted from: <i>Languages</i> 2023, 8, 82, doi:10.3390/languages8010082 | 186 |

| | |
|---|-----|
| Marlies Jansegers, Chantal Melis and Jennie Elenor Arrington Báez | |
| Diverging Grammaticalization Patterns across Spanish Varieties: The Case of <i>perdón</i> in Mexican and Peninsular Spanish | |
| Reprinted from: <i>Languages</i> 2024, 9, 13, doi:10.3390/languages9010013 | 205 |
| Delin Deng | |
| The Grammaticalization of the Discourse Marker <i>genre</i> in Swiss French | |
| Reprinted from: <i>Languages</i> 2023, 8, 28, doi:10.3390/languages8010028 | 225 |
| Shima Salameh Jiménez | |
| Paths of Constructionalization in Peninsular Spanish: The Development of “Pues Eso”. A 20th Century Case | |
| Reprinted from: <i>Languages</i> 2023, 8, 289, doi:10.3390/languages8040289 | 239 |

About the Editors

Renata Enghels

Renata Enghels (PhD in Romance Philology) is appointed as a Professor of Hispanic and General Romance Linguistics at Ghent University. She specializes in corpus research from a functional and cognitive perspective. She has a particular interest in verbal argument structures and focuses on the functional relations between formal–syntactic structures in Romance languages, especially in Spanish, the semantic and pragmatic roles of linguistic units and their morphological markings. Her PhD research, which was published as a book by Niemeyer (Mouton de Gruyter), investigated the interaction between mental ‘constructions’ of different perception modalities and the semantic and syntactic behavior of the corresponding verbs. She frequently publishes work on causative infinitive constructions, the polysemy of cognate verbs, grammaticalization, constructionalization, pragmatic markers, youth language and the nature of codeswitching.

Marlies Jansegers

Marlies Jansegers is an Assistant Professor at Ghent University and is affiliated with the Department of Linguistics. She studied Romance Philology at Ghent University and was granted a PhD in Linguistics at the same university in 2015. Her main research focus is on the semantics of perception verbs. In her PhD dissertation, she focused on the semantics of one particular so-called “minor” perception verb—the Spanish verb ‘sentir’—from a cross-linguistic (French–Spanish–Italian), synchronic and diachronic perspective. Her research has since built on this, with new avenues of research being pursued into the quantitative study of meaning in a broad sense, including not only (lexical) meaning but also discourse function, information structure, and other pragmatic phenomena such as apology markers. Methodologically, she is interested in applying corpus-based techniques to various domains such as contrastive linguistics, historical and dialectal variation and intercultural pragmatics. She is the author of “Hacia un enfoque múltiple de la polisemia. Un estudio empírico del verbo multimodal ‘sentir’ desde una perspectiva sincrónica y diacrónica” (De Gruyter 2017).

Tom Bossuyt

Tom Bossuyt is a PhD student at the Department of Linguistics at Ghent University. He studied German and English Linguistics and Literature at Ghent University (2013–2017) and General Linguistics at the University of Freiburg (2017–2020). Since November 2020, he has been working on an FWO research project, in which he investigates adverbial clauses known as concessive conditionals from a broad typological perspective.

Preface

Language is not static; it is a living, evolving entity shaped by the people who use it. The volume *Grammaticalization across Languages, Levels, and Frameworks* studies the transformative processes that turn different kinds of expressions into grammatical and pragmatic structures. It gathers insights from leading scholars who explore the nuances of grammaticalization, lexicalization, and constructionalization across a wide array of languages. Our aim is to provide a comprehensive view of how these processes operate across different linguistic systems, shedding light on the intricate relationships between form, meaning, and context. This collection further illustrates the richness of linguistic diversity and the ongoing evolution of language, with the hope that it will inspire further exploration.

Renata Enghels, Marlies Jansegers, and Tom Bossuyt

Editors



Editorial

Introduction to the Special Issue Grammaticalization across Languages, Levels, and Frameworks

Renata Enghels *, Marlies Jansegers and Tom Bossuyt

Department of Linguistics, Ghent University, 9000 Ghent, Belgium; marlies.jansegers@ugent.be (M.J.); tom.bossuyt@ugent.be (T.B.)

* Correspondence: renata.enghels@ugent.be

Over the last few decades, we have witnessed a proliferation in studies on phenomena such as grammaticalization (cf., among many others, Heine et al. 1991; Traugott and Heine 1991; Wischer and Diewald 2002; Hopper and Traugott 2003; Narrog and Heine 2011), lexicalization (among others, Brinton 2002; Hopper and Traugott 2003; Brinton and Traugott 2005), and, more recently, constructionalization (Traugott and Trousdale 2013; Hilpert 2013; Traugott 2014; Boogaart et al. 2014; Barðdal et al. 2015; and contributions in Giacalone Ramat et al. 2013). According to their most basic definitions, these notions refer to different phenomena of language change, which typically begin with a lexical form with a concrete meaning and end, respectively, in a grammatical unit or zero, a new lexical unit, or a fixed grammatical pattern or construction. Extensive bibliographies reflect ongoing discussions concerning every possible aspect of these linguistic processes, including the basic terminology and their internal relationships. Still, a common feature of grammaticalization theories is that they explain particular cases of changing grammar primarily as being a consequence of concrete language use. As a consequence, new knowledge on the nature of overall language change can best be gathered by closely studying particular phenomena in empirical data, coming from different languages, and incorporating insights from different frameworks. Therefore, this Special Issue aims at reporting on new and current interests and developments in this established yet ever-evolving field of research. It features a diverse collection of studies covering a broad spectrum of topics, methodologies, datasets, and languages. Many of these studies were initially presented at the Grammar and Corpora conference held at Ghent University in the summer of 2022.

Since the 1980s, the field of research has seen an increase in studies exploring languages with varied typological profiles and genealogical affiliations. Significant advances have been achieved in identifying common grammaticalization paths across languages (Viberg 1999), leading to the development of various clines or hierarchies based on extensive comparative data (Heine et al. 1991; Haspelmath 2004). However, the majority of these studies focus on specific linguistic phenomena within individual languages, often leveraging increasingly larger corpora. Predictably, the most substantial progress has occurred in languages with abundant historical data, primarily European languages such as those in the Romance and Germanic branches of Indo-European. Despite this focus, the field continues to diversify. For instance, the new millennium has witnessed growing research on grammaticalization in sign languages, which have traditionally been underrepresented in linguistic studies, as well as on comparative analyses of grammaticalization in both spoken and signed languages (e.g., Pfau and Steinbach 2011). The papers included in this volume include well-studied languages like Spanish (with contributions from Garachana and Sansiñena, Torres and Enghels, Salameh Jiménez, and Fuentes Rodríguez), Italian (Morei), and English (Smith). Additionally, the volume features research on less commonly studied languages such as Romanian (Illoaia) and Old Catalan (Torres-Latorre and Sentí). There are also contrastive studies between different languages (Paoli) and dialects (with Jansegers et al. focusing on Spanish and Deng on French). Moreover, the volume includes research on (German) sign language (Otte et al.), further highlighting its wide linguistic scope.

Citation: Enghels, Renata, Marlies Jansegers, and Tom Bossuyt. 2024. Introduction to the Special Issue Grammaticalization across Languages, Levels, and Frameworks. *Languages* 9: 254. <https://doi.org/10.3390/languages9070254>

Received: 15 July 2024

Accepted: 16 July 2024

Published: 21 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Next, research on grammaticalization spans various levels of linguistic analysis. Initially, studies predominantly focused on the morphosyntactic dimensions of grammaticalization, viewing it primarily as a process of morphosyntactic reduction (cf., among others, Lehmann [1982] 1995; Hopper 1991; Haspelmath 2004). Conversely, some scholars have emphasized not just the aspect of loss but also the semantic enrichment that accompanies this process (Bybee et al. 1994; Diewald 2002; Heine 2002; Traugott 2003). As a consequence, more recent approaches extend beyond morphosyntax and semantics to explore grammaticalization within the context of discourse. These studies analyze grammaticalization in real spoken conversations and examine its connection to broader cognitive processes involved in the interaction between speaker and addressee. Consequently, both the form and meaning are given equal importance, with ‘meaning’ encompassing not only lexical significance but also discourse function, information structure, and various other pragmatic aspects. Against this background, this volume explores grammaticalization across multiple levels of linguistic analysis.

Lastly, our understanding of grammaticalization processes can be enriched by incorporating new insights from relevant theoretical frameworks. The concept of ‘emerging grammar’ (Hopper 1987) (partially) explains why grammaticalization has mainly developed in parallel with the development of cognitive and functional linguistics. Indeed, besides describing the formal paths of grammaticalization, researchers have increasingly focused on the motivations driving these changes and exploring why they occur. Hence increasing attention was paid to the functional meaning of grammaticalization, incorporating various semantic and pragmatic accounts. Since the early 21st century, there have been efforts to examine diachronic change through the lens of construction grammar. Although most studies on ‘constructionalization’ acknowledge that this approach does not fundamentally deviate from traditional grammaticalization studies (see references above), they highlight its utility in providing a comprehensive framework for understanding both lexical and grammatical changes. Furthermore, the adoption of the construction concept allows for more precise measurements of the gradual nature of grammatical change, facilitating detailed investigations into its productivity. The cross-fertilization between grammaticalization and construction grammar, particularly in the notion of constructionalization, is notably present in the papers of Garachana and Sansiñena, Torres and Enghels, Illoaia, Smith, Salameh Jiménez, and Fuentes Rodríguez.

The volume builds progressively through various levels of linguistic analysis, beginning with two papers on the grammaticalization of morphological phenomena (Paoli and Otte et al.). Advancing to the constructional level, four papers focus on grammaticalization of specific complex syntactic constructions, in particular verbal periphrases (Garachana and Sansiñena, Torres Soler and Enghels, Illoaia, and Smith). Subsequently, the focus shifts to verbal tenses, thereby adding another layer of complexity (Torres-Latorre and Sentí, Morei). Ultimately, the analysis reaches the discourse level, where grammaticalization processes of pragmatic markers come into play (Jansegers et al., Deng, Salameh Jiménez, and Fuentes Rodríguez).

Sandra Paoli’s paper, titled “Gradualness of Grammaticalization and Abrupt Change Reconciled: Evidence from Microvariation in Romance”, emphasizes the interplay of grammaticalization processes that exhibit both gradual and abrupt characteristics. To this end, her paper presents an in-depth empirical analysis of clitic loss across several Romance languages, namely Brazilian Portuguese, Raeto-Romance, northeastern Italo-Romance varieties, and French. From a theoretical perspective, the concepts of gradience (variation within a category at a given time) and gradualness (the accumulation of small changes over time) are discussed, and it is further shown how localized changes (gradience) contribute to the broader evolutionary trajectories of grammatical structures (gradualness). In conclusion, the study identifies common patterns where certain cells within the clitic paradigms are more susceptible to loss, often influenced by animacy, specificity, and other grammatical features. These patterns are consistent across the studied languages (for instance, third-person and less animate referents are typically the first to exhibit loss).

However, by examining microvariation, the paper reveals that what may seem like abrupt changes on a micro level contributes to a perception of gradual change on a macro scale. Overall, this approach allows for a nuanced understanding of the mechanisms driving grammaticalization, aligning well with the Special Issue's focus on empirical data and theoretical insights in evolving research in this domain.

The paper "Numeral Incorporation as Grammaticalization? A Corpus Study on German Sign Language (DGS)" by Felicitas Otte, Anke Müller, Sabrina Wähl, and Gabriele Langer examines numeral incorporation in German Sign Language (DGS), exploring whether this phenomenon reflects a process of grammaticalization. Based on a DGS corpus, the study recurs to an apparent-time method to infer diachronic trends. In sign languages, numeral incorporation refers to the phenomenon where numeral signs merge with lexical signs (mostly units of time, measurements, or quantities) to form a new, composite sign (e.g., "three weeks"). The analysis of the corpus identifies a progression in the form of these combinations from free morphemes to cliticized morphemes and finally to bound morphemes (affixes), or a movement from independent numeral and base signs toward a unified, grammaticalized form where numeral signs become an integral part of the lexical sign. However, while the study observes some characteristics of grammaticalization, such as increased frequency and fixedness of use among younger signers, it concludes that numeral incorporation does not fully conform to traditional models of grammaticalization because certain expected features of grammaticalization, such as obligatoriness or fossilization, are absent.

In their study, "Combinatorial Productivity of Spanish Verbal Periphrases as an Indicator of Their Degree of Grammaticalization", Mar Garachana and María Sol Sansiñena explore the grammaticalization process of (near-synonymous) Spanish verbal periphrases, focusing on *dejar de* + INF and *parar de* + INF. Using a constructionist, usage-based approach, the authors investigate these periphrases' semantic areas, functional distributions over time, and collostructional patterns to discuss their productivity and how this relates to their degree of grammaticalization. The analysis of the historical data points towards a significant divergence in the evolution of *dejar de* + INF and *parar de* + INF, with the latter not paralleling the former in terms of combinatorial patterns, semantic fields, or lexical productivity. While *dejar de* + INF has reached a stable state in its grammaticalization process (as shown, for instance, by its limited combinatorial productivity), *parar de* + INF continues to evolve and extend to new contexts, showing variability in its integration into the grammatical system. Overall, the paper provides a nuanced understanding of how these two similar yet distinct periphrases have developed in Spanish, contributing to broader discussions on language change, grammaticalization, and the impact of semantic and functional factors on the evolution of language structures.

The paper "From Motion to Causation: The Diachrony of the Spanish Causative Constructions with *traer* ('bring') and *llevar* ('take')" by Julio Torres Soler and Renata Enghels investigates the historical evolution (between the 13th and 20th centuries) of specific Spanish causative constructions involving the verbs *traer* and *llevar*. The study contributes to the understanding of grammaticalization by illustrating how changes at the lexical and semantic levels impact syntactic structures and the productivity of grammatical patterns in a language, resonating with themes of diachronic change and construction grammar. In concrete, the paper identifies a specific linguistic subschema characterized by verbs of 'caused accompanied motion' that have semantically specialized to denote indirect causation. This specialization is marked by a low level of syntactic integration between the causative verb and the infinitive form it governs. Moreover, the historical semantic shifts of *llevar* and *traer* significantly influenced their integration into causative constructions. Initially, *llevar* was more associated with unbounded motion, which delayed its adoption into causative uses. From the 16th century onward, as the meanings of the verbs diverged, with *llevar* becoming more about goal-oriented motion and *traer* about motion toward the speaker or addressee, their roles in causative constructions also shifted.

Next, the paper by Mihaela Ilioiaia, “Constructing Meaning: Historical Changes in MIHI EST and HABEO Constructions in Romanian”, studies the evolution of two competing Latin patterns originally used to express possession, highlighting their different semantic and syntactic trajectories in Romanian compared to other Romance languages. Remarkably, unlike other Romance languages, where *habeo* replaces *mihi est* in possessor and experiencer contexts, Romanian preserves both constructions. However, *mihi est* is more commonly used in experiencer contexts (involving psychological and physiological states), suggesting a specialized functional evolution, whereas *habeo* tends to cover a broader range of uses, including abstract possession. Ilioiaia utilizes a diachronic corpus to track changes over time, focusing on how different state nouns are used within these constructions. The paper further evaluates whether the changes observed in these constructions can be classified under grammaticalization or constructionalization, eventually supporting the idea of constructional change. Overall, in the same line of ideas as Garachana and Sansiñena’s paper, this study emphasized the role of functional differentiation in the grammaticalization process.

In “Productivity from a Metapragmatic Perspective: Measuring the Diachronic Coverage of the Low Level Lexico-Grammatical Construction Have the N (Body Part/Attitude) to ↔ <Metapragmatic Comment> Using the COHA”, Chris A. Smith addresses the issue of constructional change and how to determine its emergence, as well as the productivity of such changes. The overarching question of the paper is how semantics affects productivity, in the generative sense of extensibility of a construction (a form-meaning pairing). The study focuses more precisely on the metapragmatic reaction to perceived insolence via the expression *have the N to* (e.g., *have the guts to*). Theoretically, it adopts the constructional network approach, implying that constructions form a network of interrelated constructions at a higher and lower level that can affect each other. Methodologically, this paper presents a lexicalist bottom-up approach to the construction and contributes to the question of how to overcome methodological issues relating to a qualitative rather than quantitative approach to constructional architecture and the relative productivity of constructions. In doing so, it relies on a fine-grained distributional semantic analysis using the Corpus of Historical American English (COHA). It shows that studying low-level lexico-grammatical patterns can provide a more reliable understanding of the emergence and development of a family of constructions and that a global view of constructional architecture at multiple levels should be pertinent to identify the extensibility potential of the construction.

The paper “Clitic Placement and the Grammaticalization of the Future and the Conditional in Old Catalan” by Aina Torres-Latorre and Andreu Sentí focuses on a well-known process of morphological change in Romance languages—the grammaticalization of Latin periphrasis *cantāre habeō* towards the Romance future and conditional tenses—but concentrates on a less-studied language, namely Catalan. The study adopts a descriptive, empirical, and quantitative approach and constitutes the first exhaustive depiction of the frequencies of the different possibilities for clitic placement (proclisis, mesocclisis, and enclisis) associated with Catalan future/conditional constructions throughout the Middle Ages and the 16th century. As such, it completes the panorama of Ibero-Romance varieties on this phenomenon and allows for a comparison as to the placement of Catalan within the Romance continuum. As it turns out, the grammaticalization of future and conditional in the 13th and 14th centuries was more advanced in the eastern languages of the Iberian Peninsula, such as Catalan, than in the western ones. The results obtained for Old Catalan appear to confirm a language contact phenomenon in which Catalan (and probably Occitan) spread the more grammaticalized future/conditional to Navarro-Aragonese and Castilian. The authors prove these significant quantitative differences through a detailed analysis of all data in a well-balanced corpus of selected texts from the Catalan historical digital corpus (CICA).

A similar focus on the grammaticalization of tense is present in the paper “Not Only Anteriority in the Past: The Functions of the Pluperfect in Spoken Italian”, where Eleonora Morei conducts a corpus-based approach on the Italian pluperfect to empirically evaluate

the hypothesis of the existence of an aoristic (i.e., deictic) use of this tense. To this end, a dataset of spoken spontaneous speech was compiled, based on the (northern) Italian ParlaTO corpus. She shows that the Italian pluperfect has developed secondary functions that can be explained by (inter)subjectification paths of grammaticalization and demonstrates how tense, aspect, and modality intertwine in defining the Pluperfect's semantics.

The use of spontaneous spoken data also characterizes the study conducted by Marlies Jansegers, Chantal Melis, and Elenor Arrington, "Diverging Grammaticalization Patterns across Spanish Varieties: The Case of *perdón* in Mexican and Peninsular Spanish". By means of a corpus-based comparative analysis based on spoken conversations and interviews in both varieties of Spanish, this study empirically investigates the contemporary grammaticalized uses of the apologetic marker *perdón* ('sorry'). Contrary to what usually happens to other pragmatic markers, the point of departure for *perdón* rests in an element that already has an illocutionary force at the onset, the original illocutionary act of regret. It is argued that this original illocutionary material is recruited for further illocutionary and discursive uses, determining the contemporary grammaticalized uses and values of *perdón*. That is, the grammaticalization process of *perdón* embodies a process that affects the illocutionary potential of the speech act itself. Therefore, rather than semantic bleaching, this can be best described as a case of progressive weakening of its illocutionary force. Interestingly, this even seems to give rise to diverging patterns of grammaticalization across varieties of the same language, since evidence from both corpora suggests a more advanced stage in the grammaticalization process of *perdón* in Mexican Spanish.

A similar process of diverging grammaticalization patterns across varieties of the same language was discovered by Delin Deng in her paper "The Grammaticalization of the Discourse Marker *genre* in Swiss French". The study focuses more precisely on the discourse marker *genre* in Swiss French. By conducting an apparent-time variationist analysis based on oral data taken from the OFROM corpus (le corpus Oral de Français de Suisse Romande), the study offers a quantitative approach to the grammaticalization status of *genre* in Swiss French and its correlation with social factors, such as age, gender, sociolinguistic situation, and socio-educational status of the speakers. It also examines phonological reduction to support the claim that *genre* in Swiss French is undergoing an independent process of grammaticalization from that in Hexagonal French documented in the literature. Moreover, the study provides new insights into the development of the same particle in different regions being conditioned by different social factors: while the ongoing change of *genre* is led by female speakers in Swiss French, it is first led by male speakers, then quickly spread to both gender groups in Hexagonal French.

The relevance of studying processes of grammaticalization and constructionalization in conversational formulas is further illustrated in the study by Shima Salameh Jiménez, "Paths of Constructionalization in Peninsular Spanish: The Development of "Pues Eso". A 20th Century Case". The study traces the recent diachronic evolution of Peninsular Spanish *pues eso* in the 20th century as a construction by analyzing linguistic patterns related to pre- and post-constructional processes, including syntactic, semantic, and pragmatic features. Methodologically, the proposal incorporates units and positions from the Val.Es.Co. model of discourse segmentation, and the data were retrieved from the CDH corpus (Corpus del Diccionario histórico de la lengua Española, Real Academia Española), with particular attention to written texts reproducing orality. By adopting a constructional approach, the study reveals that the current functions covered by *pues eso* were mainly consolidated during the 20th century, but with some important antecedents in the 16th and 17th centuries, when agreement and self-reinforcement functions seem to have originated.

Catalina Fuentes Rodríguez's paper "From Peripheral Structure to Discourse Operator: No Veas" studies the grammaticalization and cooptation processes of the Spanish expression *no veas*. This expression, originally part of more syntactically bound structures, has evolved into a more flexible discourse operator with various pragmatic functions. The study uses corpus data from both the CORPES XXI and the MEsA corpus and focuses on instances where *no veas* is used in different contexts, starting from a literal verb phrase

(with the meaning of seeing or watching something), progressively used for intensification, to a peripheral usage as a discourse operator indicating surprise or emphasis. It is argued that the evolution of *no veas* reflects cooptation, where it gains new pragmatic functions without undergoing complete grammaticalization, which would typically involve more extensive semantic bleaching and syntactic fixation. Moreover, it illustrates a process of subjectification as the expression transitions from describing external events to expressing the speaker's stance. Overall, the paper contributes to the theoretical discussion on how linguistic elements evolve from syntactic components to playing roles in discourse structuring, adapting to the communicative needs of speakers.

In conclusion, as has been showcased by the papers included in this volume, the exploration of grammaticalization has yielded substantial insights into the evolution of language at multiple levels of analysis. However, we believe that future research should increasingly focus on the social factors that facilitate or inhibit grammaticalization processes. Specifically, the role of social networks in shaping linguistic change warrants closer examination. Recent studies have expanded the scope of grammaticalization to include discourse-level phenomena, highlighting the intricate interplay between structure and meaning. Moving forward, it is imperative to integrate social dimensions into this framework. Alongside lexical, semantic, and pragmatic information, social information such as the dynamics of speaker communities, patterns of language contact, and the influence of social hierarchies could provide a more comprehensive understanding of grammaticalization. By incorporating these social factors, future research can offer a more holistic view of how grammatical structures emerge and evolve within their sociocultural contexts, thereby enriching our understanding of language change and development.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer, and Spike Gildea, eds. 2015. *Diachronic Construction Grammar*. Amsterdam: John Benjamins.
- Boogaart, Ronny, Timothy Coleman, and Gijsbert Rutten. 2014. *Extending the Scope of Construction Grammar*. Berlin and New York: Mouton De Gruyter.
- Brinton, Laurel J. 2002. Grammaticalization versus lexicalization reconsidered: On the "late" use of temporal adverbs. In *English Historical Syntax and Morphology: Selected Papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000*. Edited by Teresa Fanego, María José López-Couso and Javier Pérez-Guerra. Amsterdam and Philadelphia: John Benjamins, pp. 67–97.
- Brinton, Laurel J., and Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Bybee, Joan L., Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago Press.
- Diewald, Gabriele. 2002. A model for relevant types of contexts in grammaticalization. In *New Reflections on Grammaticalization*. Edited by Ilse Wischer and Gabriele Diewald. Amsterdam and Philadelphia: John Benjamins, pp. 103–120.
- Giacalone Ramat, Anna, Caterina Mauri, and Piera Molinelli, eds. 2013. *Synchrony and Diachrony: A Dynamic Interface*. Amsterdam: John Benjamins.
- Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In *Up and Down the Cline: The Nature of Grammaticalization*. Edited by Olga Fischer, Muriel Norde and Harry Perridon. Amsterdam and Philadelphia: John Benjamins, pp. 17–44.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In *New Reflections on Grammaticalization*. Edited by Ilse Wischer and Gabriele Diewald. Amsterdam and Philadelphia: John Benjamins, pp. 83–101.
- Heine, Bernd, Ulrike Claudi, and Friederike Hünemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago and London: University of Chicago Press.
- Hilpert, Martin. 2013. *Constructional change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.
- Hopper, Paul J. 1987. Emergent Grammar. *Berkeley Linguistics Society* 13: 139–57. [CrossRef]
- Hopper, Paul J. 1991. On Some Principles of Grammaticization. In *Approaches to Grammaticalization*. Edited by Elizabeth Closs Traugott and Bernd Heine. Amsterdam and Philadelphia: John Benjamins, pp. 17–36.
- Hopper, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization*. Munich: Lincom Europa. First published 1982.
- Narrog, Heiko, and Bernd Heine, eds. 2011. *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.

- Pfau, Roland, and Markus Steinbach. 2011. Grammaticalization in Sign Languages. In *The Oxford Handbook of Grammaticalization*. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press, pp. 683–95.
- Traugott, Elizabeth Closs. 2003. Constructions in grammaticalization. In *The Handbook of Historical Linguistics*. Edited by Brian D. Joseph and Richard D. Janda. Oxford: Blackwell, pp. 624–47.
- Traugott, Elizabeth Closs. 2014. Toward a constructional framework for research on language change. *Cognitive Linguistic Studies* 1: 3–21. [CrossRef]
- Traugott, Elizabeth Closs, and Bernd Heine. 1991. *Approaches to Grammaticalization*. Amsterdam and Philadelphia: John Benjamins.
- Traugott, Elizabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford Studies in Diachronic and Historical Linguistics. Oxford: Oxford University Press.
- Viberg, Ake. 1999. The polysemous cognates Swedish *gå* and English *go*: Universal and language-specific characteristics. *Languages in Contrast* 2: 87–113. [CrossRef]
- Wischer, Ilse, and Gabriele Diewald, eds. 2002. *New Reflections on Grammaticalization*. Amsterdam and Philadelphia: John Benjamins.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

From Peripheral Structure to Discourse Operator: *No Veas*

Catalina Fuentes Rodríguez

Department of Spanish Language, Linguistics and Theory of Literature, University of Seville, 41004 Sevilla, Spain; cfuentes@us.es

Abstract: This work describes the cooptation/grammaticalization process developed by the expression *no veas*. The hypothesis it defends, endorsed by previous research, considers that in this process, this expression appears in different constructions, originally integrated in the clause or peripheral, and is involved in a process of fixation towards pragmatic markers. The parenthetical distribution is fundamental to this subjectification and constructionalization. In some cases, the cooptation process has not ended, and constructions can act as semi-fixed patterns. These constructions provide procedural content but, at the same time, allow lexical choice in part of their structure. In other cases, we are already dealing with fixed markers. The Spanish structure *no veas* can appear as a free verbal construction (“*No veas tanto la televisión*” (*don’t watch television so much*)), as a semi-free intensification construction (“*no veas lo enfadado que está*” (you wouldn’t believe how angry he is), “*está gritando que no veas*” (he’s shouting like you wouldn’t believe)), or as a discourse operator (“*El campo está lleno, no veas*” (the stadium’s full, unbelievable)). In the latter case, the structure appears peripheral and displays a high degree of mobility (it could be inserted at the beginning of the utterance or appear in an intermediate position, not just at the end) and functions as a modal operator of surprise (a mirative) or a comment with intensifying meaning.

Keywords: construction; cooptation; discourse operator

1. Introduction

This paper focuses on the process of cooptation (Heine 2013) and grammaticalization (Traugott and Trousdale 2013) that certain Spanish constructions undergo to become ethicals (Kaltenböck et al. 2011), with procedural content and a pragmatic marker behaviour. The hypothesis it defends, endorsed by previous research, considers that in this process, the syntactic constructions were originally integrated in the clause or peripheral and have undergone a process of fixation towards pragmatic markers (e.g., *ya te digo, si quieres, si cabe, como quieras, dímelo a mí, tú sabes, aunque sea...*). As pragmatic markers, the elements provide macrostructural indications related to the stance of the interlocutor, the enunciation of the speaker, his/her subjectivity, or the informative and argumentative structure of what is being said (Fuentes Rodríguez 2003, 2017, 2018, 2022a). At the same time, some of these structures and units, such as modal operators (such as *of course, surprisingly*), may appear as a response (Padilla Herrada 2021).

The above-mentioned constructions initiate a subjectivization process, which can culminate in the grammaticalization (for others, pragmaticalization, Dostie 2004; Diewald 2011) or cooptation (Heine et al. 2021) of the structures. Hence, they indicate the continuous activity of the linguistic system and the creativity of speakers to produce expressions showing their involvement in what is said.

We will show how in some cases, the cooptation process has not ended and, following the perspective of *Construction Grammar* (Goldberg 1995, 2003; Gras and Sansiñena 2015), constructions can act as semi-fixed patterns. These constructions provide procedural content but, at the same time, allow for lexical choice in some part(s) of their structure. Other cases (e.g., *si cabe, aunque sea...*) clearly act as pragmatic markers. In this paper, the process will be illustrated by the Spanish structure (*no veas*, lit. “don’t (you) see”),

Citation: Fuentes Rodríguez, Catalina. 2023. From Peripheral Structure to Discourse Operator: *No Veas*. *Languages* 8: 254. <https://doi.org/10.3390/languages8040254>

Academic Editors: Renata Enghels, Marlies Jansengers and Tom Bossuyt

Received: 29 January 2023

Revised: 8 October 2023

Accepted: 17 October 2023

Published: 25 October 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

which can appear as a free verbal construction (“*No veas tanto la televisión*” (don’t watch television so much)), as a semi-free intensification construction (“*no veas lo enfadado que está*” (you wouldn’t believe how angry he is), “*está gritando que no veas*” (he’s shouting like you wouldn’t believe)), or as a discourse operator¹ (“*El campo está lleno, no veas*” (the stadium’s full, unbelievable)). In the latter case, the structure is peripheral and syntactically mobile (it could be inserted at the beginning of the utterance or appear in an intermediate position, not just at the end) and functions as a modal operator of surprise (i.e., a “mirative”, De Lancey 2001; Simeonova 2015; Olbertz 2012), or a comment with intensifying value.

Our methodological approach is that of pragmatically oriented linguistics, based on a macrosyntactic study² (Fuentes Rodríguez 2017) of the behaviour of a construction in an utterance and its different discursive functions (Kaltenböck 2016). Likewise, we present the constructional variation from an evolutionary perspective, in which the extrapositional nature of an element as well as the acquisition of its procedural meanings indicate the successive steps of the process that lead from free distribution to pragmatic markers, passing through semi-free constructions.

There are other similar approaches, such as the one expressed by Heine et al. (2021), for whom the evolution towards theticals (elements with a metadiscursive function related to the speaker’s subjectivity and which appear separately in the clause; see Section 2.1) is considered cooptation because the element acts with a metatextual function, although this does not imply a grammatical change or a decategorization, but an extraclausal function.³ Similarly, we consider that cooptation is a broader process, affecting more units than discourse markers, and that it can be combined with grammaticalization in a subsequent step. In addition, there are intermediate stages in which we find constructional theticals (Heine et al. 2021, p. 46). Hence, we consider it fundamental to analyse constructions such as *no veas* and their different discourse functions (macrosyntax) in discourse grammar. We differ from this position in that from our macrosyntactic approach, both parts, the clause and higher units, belong to the grammar.⁴

2. Materials and Methods

2.1. Theoretical-Methodological Premises

This study advocates a discourse syntax (or macrosyntax) approach that sustains the existence of relationships, combinatorics, and constructions that function in an utterance and between utterances, beyond the strict level of clausal dependence. The central axis of this methodological position is to distinguish between the behaviour of an element within the clause (microsyntax) and its behaviour at higher levels (sentence or text, macrosyntax). Both are part of discursive grammar (Fuentes Rodríguez 2017, 2023, to appear). The concrete form under study, *no veas*, develops a behaviour both within the clause (microsyntax) and within the utterance (as a peripheral element). Because of its specific behaviour, it is necessary to take parameters into account such as (a) the presence of constructions (conventional pairing of form and function, Goldberg 1995), (b) the distribution at the margins of the utterance, (c) the inclusion of metadiscursive procedural content that supports the textual structure, and (d) the inclusion of the communicative perspective of the participants.

A particularly interesting group of forms included within the category of discourse markers expresses procedural content linked to the formulation of the speech act (such as *frankly*), the subjectivity of the speaker (*surprisingly*), or the indications that (s)he gives to the receiver on how to interpret the message. Hence, they relate to the functions of focalization (*just*) and argumentation (*at least*). The syntax of these elements does not totally coincide with that of other connectors. Therefore, they have been called *discourse operators* (Fuentes Rodríguez 2003, 2018, 2020a). These elements function within the utterance itself and do not establish links with previous utterances, like *connectors* do. They are mobile within the utterance, occupy initial, medial, or final positions, and are not integrated into the clausal syntax. Consequently, their behaviour corresponds to that of the theticals of Kaltenböck et al. (2011), characterised as:

Theticals are elements which the speaker (or writer) presents as separate from sentence grammar in order to signal what Dik (1997, p. 396) called a “higher level orientation function”. They allow the speaker to “step out” of the confines of the linearity of communication to some extent by creating a kind of second plane of communication, not unlike “asides” on stage. This plane can be inserted spontaneously virtually anywhere and therefore lends itself particularly well to situation specific, metacommunicative information. But the plane needs to be signalled to the listener as such prosodically by separate tone units, pauses, etc., and by the suspension or loosening of constructional constraints and ensuing mobility. (Kaltenböck et al. 2011, p. 883)

In order to understand their behaviour, the authors assert that a specific method, i.e., *thetical grammar*, is required.⁵ The characteristics of theticals coincide with those indicated above.

(a) They are syntactically independent. (b) They are set off prosodically from the rest of an utterance. (c) Their meaning is “non-restrictive”. (d) They tend to be positionally mobile. (e) Their internal structure is built on principles of SG but can be elliptic. (Kaltenböck et al. 2011, p. 857)

From our perspective, connectors and operators are two paradigms of elements and not a single class of discourse markers.

2.2. *Evolutive Process*

The origin of discourse markers in general, be they connectors or operators, has been the subject of great interest and debate as, given their extra-clausal function, they do not adapt to the processes established for grammaticalization (Lehmann 1985, 2002). In fact, some authors observed that they did not suffer from semantic bleaching, as mere grammatical morphemes do; rather, they terminated with a procedural content and with a different function, in the periphery of the utterance, outside the clausal nucleus. The process in which they are involved does not lead to the total loss of semantic content and syntactic freedom but to the appearance of a new form, with another function in discourse. In order to define this process, authors have proposed terms such as *pragmaticalization* (evolution of an element that adopts pragmatic content) (Dostie 2004; Claridge and Arnovick 2010), or *discursisation* (Diewald 2011). Claridge and Arnovick define it as

the process by which a lexico-grammatical sequence or word form, in a given context, loses its propositional meaning in favour of an essentially metacommunicative, discourse interactional meaning and/or (an already pragmatic element) continues to develop further pragmatic functions or forms. (Claridge and Arnovick 2010, p. 187)

Company (2004), for her part, insists, along with researchers such as Traugott, on adopting a wider view of the process of change that includes these units. Traugott (2003, p. 645) explains grammaticalization as “the process whereby lexical material in highly constrained pragmatic and morphosyntactic context is assigned grammatical function, and once grammatical, is assigned increasingly grammatical, operator-like function”.

This process has also been said to involve a stage of fixation of the construction or constructionalization (Traugott and Trousdale 2010, 2013). Evans (2007), for instance, includes constructionalization in the evolution that gives rise to insubordinate clauses. This process involves the following steps:

A: Subordination: subordinate construction; B: Ellipsis: ellipsis of main clause;
 C: Conventionalized ellipsis: restriction of interpretation of ellipsed material;
 D: Reanalysis as main clause structure: Conventionalised main clause use of formally subordinate clauses. (Evans 2007, p. 370)

The last step is the conventionalization of the construction or “constructionalisation”: “The construction now has a specific meaning of its own and it may not be possible to

restore any ellipsed material” (Evans 2007, p. 374). This process is justified because the construction acquires new discursive functions that involve a step towards the procedural one. These new functions include:

- Indirection and interpersonal control: requests and commands, hints, warnings, and admonitions
- Modal functions of various types: epistemic and evidential meanings; deontic meanings (especially hortatives and obligation); exclamations; and evaluations
- Signalling presupposed material: negation (i.e., negative clauses have subordinate form), contrastive focus, reiteration, disagreement with assertions by previous speaker (Evans 2009, pp. 9–10)

Heine (2013, pp. 1205–6) refers to this process as cooptation, via which constructions lose their syntactic dependence of the verb, reduce their lexical content to the point of converting it into a procedural instruction, and acquire combinatorial freedom: “units such as clauses, phrases, or words are taken from the domain of sentence grammar and deployed for purposes of discourse organisation”. However, cooptation does not mean that the element has been grammaticalized. They are two processes, and grammaticalization can follow cooptation.

Traugott has also studied constructionalization in numerous works. Both underline the need to explain why the process is not equivalent to the one undergone by other elements with semantic content, such as auxiliary verbs or morphemes (e.g., the suffix-*mente* for adverbs): the element loses lexical content, combinatorial freedom, and functions as a morpheme. Here, the change moves towards a new element with procedural and functional content in the discourse (“Zero syntax, ten in pragmatics” Company 2006).

The key is in the fact that the construction displaces its semantics towards intersubjectivity (modality, enunciation, persuasion), so towards how the speaker organises the discourse. López-Couso (2010, p. 129) defines subjectification as key to grammaticalization: “the semantic-pragmatic mechanism through which meanings shift from the objective description of the external situation to the expression of the speaker’s internal perspective or attitude towards that is said”. This is a gradual process (Traugott and Trousdale 2010),⁶ which explains the different stages we can find and the constructive diversity that coexists in the discourse (layering). “The steps coincide in semantic bleaching, persistence, subjectification, decategorialisation, divergence and coalescence”, as Claridge and Arnovick (2010, p. 185) argue.⁷

2.3. Constructionalization

In our research on discourse operators, it was observed that, besides being involved in a process of constructionalization, they share their extrapositional, and thus syntactically marginal, position.⁸ Extrapositionality, in our opinion, is necessary to speak of the cooptation of a structure so that it functions as an element with procedural content (thetical). This already implies a fixation at the macrosyntactic level (metatextual, for others). There are intermediate stages in this process of constructionalization. This has been illustrated for specific items before (Fuentes Rodríguez 2014, 2020a, 2021a, 2021b, 2022b) and will be tested for *no veas* in this paper. The form *no veas* has also been observed to undergo the abovementioned fixation process and has passed through three stages, the second of which is the object of our study:

(1) Free construction: combinatorial freedom, function within the clause, and designative content. (“no veas la television”, (don’t watch television).)

(2) Peripheral function and extrapositionality: the element acquires a function outside the clause, as part of the utterance. It refers to the speaker and his or her metadiscursive activity. A determining factor in this process is that it acquires the mark of extrapositionality, appearing as an independent intonational unit placed between pauses at the left or right periphery. (*To be honest*, he prefers to live in London).

(3) The fixation of the construction as a discourse operator (*frankly, of course*): the element, now as a unit, functions within the utterance and indicates a procedural content of

modality, enunciation, focalization, or thematization, or establishes scaled argumentative indications. In this stage of fixation, the expression acts as a single element, does not permit any change in its structure, and is extrapropositional. Moreover, it acquires mobility and can be found in the medial, initial, or final position in the utterance, between pauses. We consider, then, that there has been a shift to another macrosyntactic category, that of the operators. In this integral sense of discursive grammar (Fuentes Rodríguez 2023 to appear), we can consider that there is a grammatical change.

Other studies have confirmed the process indicated above (Fuentes Rodríguez 2014, 2021a, 2022b, *para colmo, no es por nada, . . .*, among others) and allowed us to explain the genesis of the paradigm of discourse operators. These elements functioning at the level of discourse grammar or macrosyntax (Fuentes Rodríguez 2017)⁹ constantly undergo a process of renewal and illustrate ongoing language change going from the lexical to the grammatical level.

Construction Grammar (Goldberg 1995) provides us with interesting resources, above all to explain the whole process situated in step 2. It is in these intermediate phases that we witness fixation, to different degrees, along the process. The process is going to be analysed through the Spanish form *no veas*.

2.4. Materials

The study was carried out based on a search in (1) CORPES XXI, a Spanish corpus from the Real Academia Española (<https://www.rae.es/banco-de-datos/corpes-xxi>, accessed 20 December 2022) of the 21st century, and (2) the MEsA corpus, on digital discourse, compiled by C. Fuentes and her team (www.grupoapl.es/materiales-corpus/corpus-mesa, accessed on 20 December 2022). It comprises material obtained from social networks, the extension of which is indicated in the Table 1:

Table 1. MEsA corpus.

| | |
|-----------|---------------|
| Blogs | 452,499 words |
| Facebook | 293,743 words |
| Forums | 365,566 words |
| Instagram | 322,638 words |
| Websites | 537,588 words |
| Twitter | 720,584 words |
| WhatsApp | 495,769 words |
| YouTube | 398,635 words |

For the analysis of both corpora, we proceeded as follows: we carried out a lexical search for the string *no veas* in the whole corpus,¹⁰ including the variable *no veas tú*. We analysed their contexts of appearance, delimiting their distribution (peripheral or integrated in the clause) and their function in the utterance or at its margins (also as an independent element in statements or reactive interventions). We also defined their content, lexical in the cases of free constructions or procedural when acting as a pragmatic marker or semi-free construction (see Table 2).

Table 2. Results in the corpora.

| Functions | CORPES (297) | MEsA (41) |
|----------------------------------|--------------|-------------|
| Clausal verb | 43 (14.47%) | 12 (29.26%) |
| <i>No veas</i> + intensification | 165 (55.55%) | 14 (34.14%) |
| <i>Que no veas</i> | 60 (20.20%) | 6 (14.63%) |
| Discourse operator | 29 (9.76%) | 9 (21.95%) |

3. No Veas: Results and Discussion

Verbs of perception, such as *ver*, have been studied from a syntactic–semantic perspective, as they are involved in many different constructions (Horno Chéliz 2002; Alfaraz 2008). Some authors have even focused on the procedural forms they give rise to (García Miguel 2005; Montolío Durán and Unamuno 2001; Gallardo and Marín Jordá 2005; Tanghe and Jansegers 2014; Fuentes Rodríguez 2020b; Gras Manzano 2020; González Sanz 2020). Nevertheless, few studies have cited the form *no veas*. Indeed, it does not appear in dictionaries, such as the one by Briz et al. (2008, under construction) or Fuentes Rodríguez (2018).

However, *no veas* is cited as a discourse marker in the work of García Miguel (2005):

In fact, we must consider many of the intransitive uses we register with *ver*, *mirar*, *oir* and a number of other verbs of perception as discourse markers. It involves formulas such as *ya veo*, *ya ves*, *ya veremos*, *¿ves?*, *¿viste?*, *no veas*, *vamos a ver*, *mira*, *oiga*, . . . which guide the development of oral discourse and speaker–hearer interaction. (García Miguel 2005, p. 185)

This structure, composed of negation + verb, in the second person, referring to the addressee, operates in Spanish as a clausal nucleus, a free structure with a complete, meaningful exhortative content, in contexts such as the following (context 1):

1. User 23 (man): No seas tan cotilla y *no veas* el Facebook (Corpus MEsA, FB 2017 June MAS 01¹¹).

“User 23 (man): Don’t be such a gossip and don’t look at Facebook”

At the opposite end of the continuum, we find the form *no veas* used as an element that is already attached (context 3 as specified in Section 2.3), forming an independent prosodic group, being syntactically mobile, without a function within the clause, and having procedural content.

2. *No veas*, por poco no acabamos con el betadine de todo el hospital. (CORPES XXI,¹² Aranda Ruiz, P. 2003, *La otra ciudad*).

“Unbelievable,¹³ we were this close to running out of betadine in the hospital”.

The most interesting ones are the intermediate steps (point 2 in Section 2.3), where different steps of the process of change can be directly observed.

The analysis of both corpora reveals that the process of fixation is not yet complete and that there is a coexistence of different stages (layering) of constructionalization.

3.1. [No Veas + Intensification]

Another context in which *no veas* adopts a unique, procedural content is related to intensification. In this case, it is integrated into the clausal structure: “*no veas* lo que se preocupa. *No veas* lo serio que está. *No veas* la de gente/la cantidad de gente, la gente que ha venido” (you wouldn’t believe how he worries. You wouldn’t believe how serious he is. You wouldn’t believe how many people there are, the people who have come). It corresponds to the construction: [*no veas* + defined nominal group, + *cómo/qué/cuánto*. . .]. The verb complement, an object of the action *ver*, changes from being an object/stimulus of the visual perception to being a focalized and intensified reality. This semantic extension illustrates an initial process of abstraction, necessary in the process of fixation. In this use, *no veas* appears in the following settings:

(A) *no veas* + nominal group with a definite article: it expresses intensification (quantification) of and focalization on the referent expressed by the noun:

3. Y Noé se lo creyó, y se puso a construir un barco en medio del desierto. *No veas* el cachondeo de los vecinos (Expedientes X. La Biblia». *El club de la Comedia* 2001)

“And Noah believed him, and started building a boat in the middle of the desert. You wouldn’t believe how (hard) the neighbours started to joke around.”

(B) With a substantiated relative clause:

4. ¿Dónde estábamos? . . . Sí, conque un día dan el anuncio de que el rey va a hacer una gran fiesta para buscarle una novia al príncipe, que ya le tocaba casarse. . . ya estaba madurito para la cosa. . . y nada. Y *no veas* la que arman la madrastra y sus hijas. Esta es la nuestra, dicen. (2001 Sanchis Sinisterra, J. *Sangre lunar*)

“Where were we? Yes, one day they announce that the king is going to have a great party to find a bride for the prince, it was time he got married. . . he was getting too old for that. . . and nothing. You wouldn’t believe the fuss from the stepmother and her daughters. She is ours, they say”.

On some occasions, it elides the intensified structure because it can be inferred from the context. Thus, we can make a supposition in relation to the following fragment found in (5): “*No veas* ayer lo que le hizo (pasó) a tu hermano” (you wouldn’t believe what he did to your brother yesterday), always presupposing something very strange, unimaginable, surprising, or unexpected.

5. *-No veas* ayer tu hermano, en los futbolines. (2003 Aranda Ruiz, P. *La otra ciudad*)
“You wouldn’t believe your brother yesterday playing table football”

In this regard, we can observe the mirative, almost exclamative nature of *no veas*. Olbertz (2012) has argued that both concepts, namely mirativity and exclamation, are generally very closely connected but concludes that in the case of exclamation, we are looking at an illocutionary value, a modality that affects the whole sentence and expresses an emotion on behalf of the speaker:

Exclamative: the speaker expresses his/her affective stance about the propositional content evoked by the communicated content, implicating that some property or relation contained in the proposition obtains to a high degree. (Olbertz 2012, p. 91)

Compared to this, miratives point to the unexpected element of the situation: “Mir(ative) indicates that the propositional content does not coincide with the expectations of its source, i.e., the speaker or another person.” (Olbertz 2012, p. 94). So Olbertz (2012, p. 95) concludes:

there are at least three reasons for assuming that mirative and exclamative are different concepts: first, unlike exclamative illocution, mirative propositional contents can have negative polarity; secondly, mirative propositional contents can occur in non-restrictive relative clauses, which exclamative illocution cannot; finally, mirative propositional contents may occur within acts with declarative or interrogative illocution, which proves that they cannot be an illocution themselves.

For Simeonova (2015, p. 1):

Mirativity has been defined as encoding the speaker’s surprise, unprepared mind, discovery of state of affairs that is unexpected, Slobin and Aksu (1982); De Lancey (1997, 2012); Aikhenvald (2012). Mirative sentences can involve one of the following: exclamative intonation (1), some lexical expression (2), grammaticalised dedicated particle (3), or grammaticalised non-dedicated particle (4).

De Lancey (2001, p. 4) admits that mirativity, evidentiality, and modality are marked cases, and can exchange forms. Simeonova (2015, p. 4) differentiates the evidential from the mirative in the sense that “in the evidential case, the proposition is contained in the set of expectations of the speaker, while in the mirative one it is not”.

The form *no veas* alludes precisely to the unexpected, either from a qualitative or a quantitative perspective. It can be assumed that there is also a certain exclamative value behind it because the speaker is clearly involved in what is said; it is an expression of his or her subjectivity (6–8 below). Furthermore, we could include it in an illocutionary question act: “¿No veas la que armaron?” (You can’t believe what they’ve done). In this case, it could be admitted when it is a citative question: “¿has dicho no veas la que armaron?” (Did you say don’t believe what they did?).

However, semantically speaking, *no veas* implies something unexpected, spectacular, remarkable, be it in a positive or negative sense. It can refer to something unexplainable, grandiose, or unbearable. The context determines the polarity and direction of the argumentation. It involves an argumentative intensification as well as an assessment

(C) The clearest example is when *no veas* appears with a marker of intensification, with *qué, cómo, cuánto*, in general exclamative contexts:

6. De albañil trabaja desde que llegó de Tetuán, y allí no había tocado ni un ladrillo, pero... *¡no veas* tú cómo se queda con todas las coplas! Y, además... *¡qué tipazo tiene!*, *¡qué tipazo!* (2001 Naveros, M. *Al calor del día*)
 “He’s worked as a bricklayer since he got from Tetuán, and he hadn’t even touched a brick there, but... *You wouldn’t believe* how he gets with all the coplas! And, what’s more... What a body! What a body!”
7. Fran me dará varias palmadas fuertes en la espalda (hace pesas, me hundirá los omoplatos) y mientras tratará de animarme con un *No veas* cuánto lo siento tío, esto es ley de vida, no se libra nadie. (Cebrián, M. 2004, “*Aluminosis*”. *El malestar al alcance de todos*)
 “Fran will give me a number of hard pats on the back (he does weights, he’ll sink my shoulder blades) and meanwhile he’ll try and encourage me with a *You wouldn’t believe* how sorry I am mate, this is the law of life, nobody escapes.”
8. *No veas* qué saque tiene la elementa, lo que había cocinado no dio para los tres. Tuve que improvisar una ensalada y unos aperitivos. (Cebrián, M. 2004, “*Tempus fugit*”. *El malestar al alcance de todos*)
 “*You wouldn’t believe* what an appetite she has, what I had cooked wasn’t enough for the three of us. I had to make a salad and some snacks all of the sudden.”

Of these collocations, *cuánto* most clearly leans towards quantitative intensification. It places quantity at the highest point on the argumentative scale. With *cómo* or *qué*, the intensification is situated on the higher part of the assessment. *No veas* focalises the assessment and, in turn, intensifies these intensifiers. In example (6), *no veas* highlights the mode or type of service the woman had. The speaker is surprised when hearing the coplas and when seeing the amount of food eaten by the women (*qué saque tiene* (what an appetite she has)). In these contexts with exclamatives, *no veas* acts as a focuser with interactive value, calling upon the attention of the receiver of the messages. Can this be defined as mirativity, or does it only refer to the speaker?

Let us compare “what an appetite this rascal has!” (*¡qué saque tiene la elementa!*) with “you wouldn’t believe what an appetite this rascal has” (*no veas qué saque tiene la elementa*). In both cases *qué* acts as an intensifier and expresses surprise on behalf of the speaker. The utterance with *no veas* involves, furthermore, the recipient, encourages him/her to be surprised, and as such, wants to share modality.

Furthermore, when *no veas* intensifies an action, it is followed by *si*:

9. Mi madre me dijo que lo mejor era llevar siempre los churros al aire en un cordel, que metidos en esa bolsa de papel se recocían, y ya *no veas* si la bolsa era de plástico (2019 Pérez Andújar, J. *La noche fenomenal*)
 “My mother told me it was best to always carry the churros exposed to the air tied in a string, in a bag they became soggy, and then if it were a plastic bag then *you wouldn’t believe it*.”

The utterance “If it were a plastic bag” (*Si la bolsa era de plástico*) refers to a hypothetical circumstance. If this circumstance is given, “*you wouldn’t believe* (it is remarkable) how soggy they became if the bag was plastic” (*no veas* lo recocidos que se ponían si la bolsa era de plástico). From this, it can be concluded that *no veas* acts as an intensification of content that is implied, not expressed.

Moreover, *no veas* can intensify a process expressed by a gerund (10). The speaker expresses his/her surprise related to the degree of suffering when working in the fog.

10. – *No veas* currando ahí con la calina que hace ¿no Rai? (Soler, A. 2018, A. Sur)
 “Unbelievable working there in this fog, right, Rai?”

In another context (11), the predication comes first as a topic, and the intensification by *no veas* is postponed, as if it were a coda. It then appears peripherally, between pauses, as a modal commentary.

11. 8 February 2017 21:30:25: M1: Es la canción más estúpida del mundo pero me encanta!!! “It’s the most stupid song in the world but I love it!!!”;
 8 February 2017 21:30:30: M1: <audio omitido> “audio omitted”;
 8 February 2017 21:30:47: M1: Estoy viendo Tarde para la ira “I’m watching The Fury of a Patient Man”;
 8 February 2017 21:31:23: H1: que canción es esa “What song’s that”;
 8 February 2017 21:31:28: H1: tarde para la epicidad “late for epicness”;
 8 February 2017 21:36:05: M1: Al actor no le gusta mucho Dani Rovira no “The actor doesn’t like Dani Rovira much”;
 8 February 2017 21:36:16: M1: Porque lo que le dijo en los Goya *no veas* “Because you wouldn’t believe what he said to him at the Goyas”;
 8 February 2017 21:36:23: M1: No estaba mucho para el humor “He wasn’t really in the mood for humour” (Corpus MEsA, WA 2017 ene–jun).

We have even collected a number of examples where *no veas* appears without any complement. The intensifying meaning can be inferred from the context:

12. La vida es lo que es, precaria y penosa se la mire por donde se la mire. Y si encima, se tienen pájaros en la cabeza o más ideas y pensamientos de los necesarios, entonces ni te cuento. Luego viene la muerte, y *no veas*. No te voy a meter miedo pero sentado aquí, a tu lado, se oye silbar el filo de la guadaña, lo que indica que estás más maduro de lo que quisieras. (Díez, L.M. 2002, *El oscurecer (Un encuentro)*)
 “Life is what it is, uncertain and painful whatever way you look at it. And if on top of it if they’ve got their heads in the clouds or more ideas or thoughts than necessary, then forget it. Then comes death, and *you wouldn’t believe it*. I’m not going to scare you but sitting here, by your side, you can hear the whistle of the blade of the scythe, which means you’re older than you’d like.”

In fragment (13), the use of *no veas*, inserted in a parenthetical utterance, also presupposes an intensified content:

13. –Eso mismo pensé yo, Vicky; pero me dije: “Susi, hija, ya que has pagado el gimnasio y te has comprado los calentadores y la malla (que *no veas* tú para encontrarla de mi talla), no te vas a echar atrás ahora por un simple ‘estiramiento del muslo’”. (2005 Bodega Estévez, L. . . [et al.]: *La maruja liberá*)
 “That’s what I thought, Vicky; but I told myself: “Susi, babe, now that you’ve paid for the gym and you’ve bought the legwarmers and leggings (and now *you wouldn’t believe it* to find my size), you’re not going to back out because of a simple ‘muscle strain’”.

It is equivalent to “you wouldn’t believe how difficult it was to find it, what I had to do to find it” (que *no veas* tú lo difícil que ha sido encontrarla, lo que tuve que hacer para encontrarla). In this context, the second-person pronoun is made explicit, as an additional means of intensification.

3.2. [Que No Veas]

Besides *no veas*, another intensifying construction is the one that appears as an intensifying coda (*que no veas*) after the verb or nominal group. *Que no veas* intensifies the action or referent expressed by the noun, while at the same time indicating a subjective burden by the speaker. The intensification has an expressive meaning:

14. Porque con los zapatos que te compraste, te huelen los pies en la noche *que no veas*. (Salcedo, H. 2002, *Obras en un acto*)

“Because with the shoes you bought, *you wouldn’t believe* how your feet smell at night”.

In *Te huelen los pies que no veas* (you wouldn’t believe how your feet smell), the construction *que no veas* acts as a complement of intensification of the action: *te huelen muchísimo* (they smell so much). The original structure could be a consecutive sentence in which (elided) quantification has been replaced by the point of reference presented by *that [que]*: “your feet smell (so much) that you wouldn’t believe [how unbearable] it is [te huelen los pies (tanto) que no veas (lo insoportable que es)]”.

This degree of intensification of the verbal action is found in other cases of *que no veas*, as a complement of an action.

15. Contigo ha tenido ella siempre mucha confianza y te quiere *que no veas*, yo creo que tanto como nosotras; por eso pensé que a lo mejor sabías algo. (Salvador Caja, G. 2002, *El eje del compás*)

“She has always had a lot of trust in you and *you wouldn’t believe* how much she loves you, I think as much as us, that’s why I thought maybe you’d know something.”

16. - Y lo mío ha sido mucho más difícil, no sé si lo sabes...(.) Que el que iba a por ti era gordo, pero los míos corrían *que no veas*. Mira si he corrido, que me he secado con la carrera. (Casavella, F. 2002, *Los juegos feroces*)

“And mine has been much more difficult, I don’t know if you know...(.) The one who went for you was fat, but *you wouldn’t believe* how mine ran. Did I run or what, I dried out running”

In example (15), the utterance “it’s unbelievable how much he loves you” (*te quiere que no veas*) is equivalent to “he loves you so much” (*te quiere muchísimo*); in (16), “it’s unbelievable how they ran” (*corrían que no veas*) is equivalent to “they ran very fast” (*corrían mucho*).

However, with other verbs, intensification does not just consist of an elevated degree of quantification (corresponding to *mucho* “a lot”). In other contexts, it involves a particular mode, as in *ponerse que no veas* (to get/become like you wouldn’t believe) (17). It implies the use of “act like a frenzied person, become angry” (*ponerse como un energúmeno, enfadarse*). In example (17), rather than indicating the emotional state the person acquires, it indicates an intensified evaluation.

17. Lo que me ha dejado hundida ha sido lo del libro. No tenías que haberlo aceptado, Mariate, pero a ver qué podía hacer yo, porque él también se ha puesto *que no veas*, (Antolín, E. 2005, *Final feliz*)

“What really got me down is the thing about the book. You didn’t have to accept it, Mariate, but it would have been difficult for me to do anything, because he’s also become like *you wouldn’t believe*”.

With nouns, *no veas que* is frequent in indefinite noun phrases. *Que no veas* completes the indefinite referent: “de tal modo, tan grande, tan intensa *que no veas*” (in such a way, so big, so intense that you wouldn’t believe).

18. Tenías razón (una vez más, y ¿cuándo no la tienes?, me pregunto), esto del ordenador portátil es una gozada, yo al menos le he cogido un gusto *que no veas* (Aramburu, F. 2006, *Informe desde Creta*)

“You were right (again, and when aren’t you? I ask myself), all this with the laptop is fantastic, I for once enjoy it *like you wouldn’t believe*.”

We also find it with *cada*, intensifying “santa soquetiza”:

19. - Él fue mi maestro porque cuando yo era chico en la escuela me ponían *cada santa soquetiza que no veas*. (Esquivel, L. 2001, *Tan veloz como el deseo*)

“He was my teacher because when I was little at school they punched the hell out of me *like you wouldn’t believe*.”

3.3. [No Veas] as an Independent Element

Finally, *no veas* can appear as a peripheral (exclamative) comment, that constitutes an independent utterance. Rather than being integrated into the clause, it appears between pauses. It establishes an exclamative assertion of surprise by the speaker, anteposed or postposed to the content it wants to highlight, and explicitly includes the receiver of the message in the communicative act. As such, it is the equivalent of: "I call the attention of the receiver regarding what has been said + I find it surprising, astonishing + I intensify or evaluate something as unexpected".

It can refer to an utterance that follows (when *no veas* is in the initial position) or precedes (when placed in the final position). It can be a response to an intensified comment or an evaluation at an elevated degree (20).

20. es que esto de justificarse por todo es que me mata
por todo/por todo
sí
no veas (2002 CORALES *La ventana: entrevista a Luz Casal, 29/11/02*)
"it is that this thing of justifying yourself for everything kills me
for everything/for everything
yes
unbelievable."

In example (20), *no veas*, being a modal operator, constitutes a turn in itself. In example (21), the modal discourse operator *no veas* is followed by the argument that justifies the evaluation.

21. -Cómo ha ido la mañana -preguntó uno.
-*No veas*, por poco no acabamos con el betadine de todo el hospital. (Aranda Ruiz, P. 2003, *La otra ciudad*)
"-How did the morning go—someone asked
-*Unbelievable*, we were this close to running out of betadine in the hospital".

In the final position, *no veas* acts as a peripheral complement in the coda and as such, expresses an intensifying modal comment on what precedes (Fuentes Rodríguez 2012):

22. Es una señora y de guapa, *no veas*. (...) Es guapa de verdad...¡ (Paz Pasamar, P. 2004, *Historias Béticas*)
"She's a lady and *you wouldn't believe* how beautiful.(...) She's really beautiful".
23. El jefe les había echado una bulla, *no veas*, Ricardo, un broncazo de tres pares de cojones. (Correa, J.L. 2004, *Muerte en abril*)
"The boss had a right go at them, *unbelievable*, Ricardo, a real good telling off."
24. sí fue un flechazo/*no veas* (2009 PRESEGAL SCOM_M22_019)
"It was love at first sight/*unbelievable.*"
25. -No, soy de Soria. Fría de cojones, chaval. *No veas*. (Villacís, J. 2016, *El hombre de la maleta vacía*)
"No, I'm from Soria. Fucking cold, lad. *Unbelievable.*"

In (25), *no veas* intensifies the preceding expression, *de cojones*, with a final comment being an independent utterance.

In (26), *no veas* is anteposed to *hay que ser cabrones*, to which it adds a meaning of intensified evaluation.

26. -Es que *no veas*, hay que ser cabrones para ofrecerle cuatro duros a la viuda, que en realidad le corresponden, a cambio del silencio... con el cuerpo de Gumersindo aún caliente, joder. (Mestre, J. 2011, *Komatsu PC-340*)
"*Unbelievable*, they've got to be right bastards to offer the widow a pittance, which is what she's owed, really, in exchange for silence... with Gumersindo's still warm body, damn."

The operator *no veas* can also be followed by another exclamative statement, and as such, comments on what is surprising or astonishing. This is the case with *vaya tela de nombre que es Bartolo* in (27).

27. – *No veas*, Bartolo, *vaya tela de nombre*. (Soler, A. 2018, *Sur*)
 “Unbelievable, Bartolo, what a name”

In these contexts, which are quite frequent in our corpus, *no veas* expresses surprise and has a mirative function, wanting to draw the attention of the recipient and to share the emotion. From a prosodic point of view, with this function, *no veas* appears in the left or right periphery of the turn.

28. Cuando el Viejo les estaba sacando los ojos con el destornillador chillaban como bestias. Y *no veas*, al final les decíamos: “A cantar, a cantar”, y cantaban por peteneras. (Lejarza, M. and Rueda, F. 2019, *Yo confieso*)
 “When the old man was taking out their eyes with the screwdriver they screeched like wild animals. And *you wouldn’t believe it*, at the end we told them: “Sing, sing”, and they sang flamenco”

In final position, as in (29), *no veas* can express a conclusion. It is placed after the part of the discourse that is responsible for the exclamation and act of surprise.

29. Lo escayolaron y me lo llevé al circuito. Luego para volver, *no veas*. Con la escayola el Trompa no podía conducir, se lo tuvo que bajar el que venía conmigo de mecánico. (Corazón Rural, A. 2019, *Jot Down*)
 “They put it in a plaster cast and I took him to the circuit. Later to come back, *you wouldn’t believe it*. The Trunk couldn’t drive with the cast, the one I had brought with me as a mechanic had to get it out.”

This syntactic mobility is typical of discourse operators. In addition, note that the form of *no veas* is not completely fixed, given that in some cases it appears with *tú* (*no veas tú*) as a more intensified variant. The comment appears in postposition, and *no veas* functions as a pragmatic marker that values and intensifies the already intensified, with anaphoric reference to what is mentioned before it.

30. ¿qué pasa? que mucha gente se iba/a a Lugo//y de vez en cuando/y cuando venían sobre todo los hombres/ya venían cargaditos ya/*no veas tú* ya traían el vino no sé cuántas mmm copas ya. (2009 PRESEGAL SCOM_M22_019)
 “What’s happening? lots of people were going /to Lugo//and from time to time/and when above all the men/were now loaded up/*you wouldn’t believe it* they were bringing the wine I don’t know how many glasses.”

4. Conclusions

The analysis of the construction *no veas* has illustrated the recurrent process of discourse operator creation. This evolution can be described in terms of constructionalisation, which begins from a free combination of lexical elements and evolves towards a more fixed unit. This unit develops procedural content expressing intersubjectivity, including several discursive, modal, and argumentative meanings. The element is mobile and affects the entire utterance. It appears as a peripheral, extrapositional, and prosodically independent element. More specifically, *no veas* has evolved from an exhortative structure to develop an intensified mirative meaning. It directly involves the addressee, who is required to share a modal reaction of surprise.

First, in the construction [*No veas* + defined NP, + *cómo/qué/cuánto*. . .], *no veas* is still integrated into the utterance. It acquires a mirative value and intensifies content. Second, [*que no veas*] appears in the coda of the utterance and acts as a complement with intensifying functions. Both are semi-fixed constructions, which constitute an intermediate stage in the evolutive process towards discourse operators. In this last phase indeed, *no veas* operates as a single unit, appears between pauses in the periphery of the turn, and expresses mirative-intensifying procedural content.

As we have seen with *no veas*, it is necessary to take into account syntactic, semantic, and prosodic aspects in order to describe the different steps in the process of creating discourse markers. We therefore argue in favour of a macrosyntactic approach. From this perspective, it is very helpful to take into account the different types of constructions (free, semi-free, and fixed) as well as the concept of extrapropositionality, which is key in the development of discourse operators.

Funding: This research was funded by Ministerio de Ciencia e Innovación and FEDER funds, grant number “PID2021-122115NB-I00”, “Las relaciones en la construcción del discurso: un enfoque multidimensional”, and Junta de Andalucía, grant number P18-FR-2619, “Macrosintaxis del discurso persuasivo”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Corpes XXI (<https://www.rae.es/banco-de-datos/corpes-xxi>, accessed on 24 November 2022) and Corpus MEsA (<http://www.grupoapl.es/materiales-corpus/corpus-mesa>, accessed on 24 November 2022).

Conflicts of Interest: The author declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Notes

¹ For a description of what in this paper is considered a discourse operator, see Section 2.1 below.

² The microsyntax-macrosyntax distinction was established by Berrendonner (1990, 2002, 2003) and Blanche-Benveniste (2002, 2003), and corresponds to the clause syntax-discourse syntax separation. The authors advocate a discourse grammar along the same line as Kaltenböck et al. (2011), who include two parts in it: thetical grammar and sentence grammar. Both constitute discourse grammar. An approach from both postures can be seen in the 2016 monographic edition of *Modèles Linguistiques*.

³ See “grammatical changes typically associated to cooptation” in Heine et al. (2021, p. 28).

⁴ For Heine et al. (2021, p. 37) “grammaticalization neither preceded nor coincided with cooptation; rather, it must have set in subsequently”. However, the following paradox arises for the authors: “Decategorialization applies only to the internal structure of DMs whereas their external structure is shaped by cooptation, which somehow has the opposite effects of decategorialization” (p. 38). From our point of view, the problem lies in limiting grammar to the clause. The shift from functioning inside the clause to outside the clause is considered grammatical by these authors (p. 28).

⁵ Fraser (1996) also separated “discourse markers” from pragmatic markers, although he considered the former to be a subtype of the latter.

⁶ This is followed by proposals such as the one by Haspelmath (2001, p. 16539), “since grammaticalisation is generally regarded as a gradual diachronic process, it is expected that the resulting words from a gradient from full content words to clear function words”.

⁷ Claridge and Arnovick (2010, p. 185) recognise, however, differences: “Pragmatic items are not paradigmaticalised in so far as they do not join in grammatical paradigm (Cf. Brinton 1996) “(...) pragmatic items exhibit scope extension and positional freedom, whereas grammatical items show scope condensation and largely fixed syntactic positions””.

⁸ Thus, Stein and Wright (1995), Adamson (2000), and even Traugott and Dasher (2002), affirm the relationship between “subjectification” and “leftmost position in the phrase”. Nuyts (2012) links subjectivity and intersubjectivity to a dimension of individual or shared responsibility. For Company (2004, p. 2) evolution to *discourse marker* goes through “impoverishment or syntactic cancelation”, and isolation between pauses is a frequent feature in the formation of new discourse markers.

⁹ For Heine this illustrates thetical, rather than traditional sentence grammar.

¹⁰ In CORPES XXI, data have not been filtered by textual type, although most cases correspond to fiction and journalistic texts. The examples from CORPES that represent digital discourse only occupy 2.47% of the total. In the Table 2 we have indicated the absolute frequencies.

¹¹ FB: Facebook.

¹² Given that the majority of the examples come from CORPES XXI, in what follows we will only explicitly provide the reference of the texts involved in the Corpus MEsA.

¹³ Literally “don’t see (look at)”.

References

- Adamson, Sylvia. 2000. A lovely little example: Word order options and category shift in the premodifying string. In *Pathways of Change: Grammaticalization in English*. Edited by Olga Fischer, Anette Rosenbach and Dieter Stein. Amsterdam: John Benjamins, pp. 39–66.
- Aikhenvald, Alexandra. 2012. The essence of mirativity. *Linguistic Typology* 16: 435–85. [CrossRef]
- Alfaraz, Gabriela G. 2008. A look-see at the Spanish verbs of visual perception *ver* and *mirar*. *Southwest Journal of Linguistics* 27: 17–42.
- Berrendonner, Alain. 1990. Pour une macro-syntaxe. In *Données orales et théories linguistiques. Travaux de Linguistique*. Edited by Dominique Willems. Paris: Louvain Duculot, pp. 25–36.
- Berrendonner, Alain. 2002. Morpho-syntaxe, pragma-syntaxe, et ambivalences sémantiques. In *Colloque International. Macro-Syntaxe et Macro-Semantique*. Edited by Leth Andersen Hanne and Nølke Henning. Berna: Peter Lang, pp. 23–41.
- Berrendonner, Alain. 2003. Eléments pour une macro-syntaxe. Actions communicative, types de clauses, structures périodiques. In *Macro-Syntaxe et Pragmatique. L'analyse Linguistique de l'oral*. Edited by Antonietta Scarano. Roma: Bulzoni Editore, pp. 93–109.
- Blanche-Benveniste, Claire. 2002. Macro-syntaxe et micro-syntaxe: Les dispositifs de la rection verbale. In *Colloque International. Macro-Syntaxe et Macro-Semantique*. Edited by Leth Andersen Hanne and Nølke Henning. Berna: Peter Lang, pp. 95–118.
- Blanche-Benveniste, Claire. 2003. Le recouvrement de la syntaxe et de la macro-syntaxe. In *Macro-syntaxe et pragmatique. L'analyse linguistique de l'oral*. Edited by Antonietta Scarano. Roma: Bulzoni Editore, pp. 53–75. [CrossRef]
- Brinton, Laurel. 1996. *Pragmatic Markers in English. Syntactic Origins and Pragmatic Development*. Cambridge: Cambridge University Press.
- Briz, Antonio, Pons Salvador, and Portolés José. 2008. *Diccionario de partículas discursivas del español*. Available online: www.dpde.es (accessed on 27 December 2022).
- Claridge, Claudia, and Leslie Arnovick. 2010. Pragmaticalisation and Discursification. In *Historical Pragmatics*. Edited by Andreas Jucker and Irma Taavitsainen. Berlin: De Gruyter Mouton, pp. 165–92.
- Company, Concepción. 2004. Gramaticalización por subjetivización como prescindibilidad de la sintaxis. *Nueva Revista Filología Hispánica* 52: 1–27. [CrossRef]
- Company, Concepción. 2006. Zero in Syntax, Ten in Pragmatics: Subjectification as Syntactic Cancellation. In *Subjectification: Various Paths to Subjectivity*. Edited by Angeliki Athanasiadou, Costas Canalis and Bert Cornillie. Berlin and New York: Mouton de Gruyter, pp. 375–97.
- De Lancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1: 33–52.
- De Lancey, Scott. 2001. The mirative and evidentiality. *Journal of Pragmatics* 33: 369–82. [CrossRef]
- De Lancey, Scott. 2012. Still mirative after all these years. *Linguistic Typology* 16: 529–64. [CrossRef]
- Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49: 365–90. [CrossRef]
- Dik, Simon C. 1997. *The Theory of Functional Grammar II. Complex and Derived Constructions*. Berlin and New York: Mouton de Gruyter.
- Dostie, Gaétane. 2004. *Pragmaticalisation et Marqueurs Discursifs*. Bruxelles: De Boeck-Duculot.
- Evans, Nicholas. 2007. Insubordination and its uses. In *Finiteness*. Edited by Irina Nikolaeva. Oxford: Oxford University Press, pp. 366–431.
- Evans, Nicholas. 2009. Insubordination and the grammaticalisation of interactive presuppositions. In *Methodologies in Determining Morphosyntactic Change, Organized*. Osaka: Museum of Ethnography.
- Fraser, Bruce. 1996. Pragmatic Markers. *Pragmatics* 6: 167–90. [CrossRef]
- Fuentes Rodríguez, Catalina. 2003. Operador/conector, un criterio para la sintaxis discursiva. *RILCE* 19: 61–85. [CrossRef]
- Fuentes Rodríguez, Catalina. 2012. El margen derecho del enunciado. *Revista Española de Lingüística* 42: 63–93. [CrossRef]
- Fuentes Rodríguez, Catalina. 2014. Comment Clauses and the Emergence of New Discourse Markers: Spanish *lo que es más*. *Journal of Pragmatics* 61: 103–19. [CrossRef]
- Fuentes Rodríguez, Catalina. 2017. Macrosintaxis y lingüística pragmática. *Círculo de lingüística aplicada a la comunicación* 71: 5–34. [CrossRef]
- Fuentes Rodríguez, Catalina. 2018. *Diccionario de conectores y operadores del español*, 2nd ed. Madrid: Arco Libros. First Published in 2009.
- Fuentes Rodríguez, Catalina, ed. 2020a. *Operadores en proceso*. Munchen: Lincom.
- Fuentes Rodríguez, Catalina. 2020b. *Mira/mira que: Construcciones, contexto argumentativo y funciones relacionales*. *Rilce: Revista de Filología Hispánica* 36: 941–66. [CrossRef]
- Fuentes Rodríguez, Catalina. 2021b. *Ni qué hablar/ni qué decir, ¿construcciones u operadores escalares?* *Pragmalingüística* 29: 149–72. [CrossRef]
- Fuentes Rodríguez, Catalina. 2021a. *No es por nada: Spanish argumentative preface and discourse operator*. *Journal of Pragmatics* 186: 236–50. [CrossRef]
- Fuentes Rodríguez, Catalina. 2022b. Expresiones malsonantes y contenido procedimental. In *Estudios de lingüística hispánica. Teorías, datos, contextos y aplicaciones: Una introducción crítica*. Edited by Laura Mariottini and Mónica Palmerini. Madrid: Dykinson, pp. 332–61.
- Fuentes Rodríguez, Catalina. 2022a. La construcción y su rentabilidad en el ámbito de la macrosintaxis. In *El dinamismo del sistema lingüístico: Operadores y construcciones del español*. Edited by Catalina Fuentes Rodríguez, María Soledad Padilla Herrada and Víctor Pérez Béjar. Sevilla: Editorial Universidad de Sevilla, pp. 293–320.

- Gallardo, Pauls, and María Josep Marín Jordá. 2005. Marcadores discursivos procedentes de verbos perceptivos en el discurso afásico. *Revista de Investigación Lingüística* 8: 53–94.
- García Miguel, José María. 2005. Aproximación empírica a la interacción de verbos y esquemas construccionales, ejemplificada con los verbos de percepción. *ELUA* 19: 169–91. [CrossRef]
- Goldberg, Adele. E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7: 219–24. [CrossRef]
- González Sanz, Marina. 2020. Subjetividad e intersubjetividad: Un estudio sobre construcciones en proceso procedentes del verbo *ver*. In *Operadores en proceso*. Edited by Catalina Fuentes Rodríguez. Munchen: Lincom, pp. 33–71.
- Gras Manzano, Pedro. 2020. De partículas discursivas a construcciones semiesquemáticas: Construcciones introducidas por “a ver si” en español peninsular. *LEA: Lingüística Española Actual* 42: 87–107.
- Gras, Pedro, and María Sol Sansiñena. 2015. An interactional account of discourse- connective *que*-constructions in Spanish. *Text&Talk* 35: 505–29.
- Haspelmath, Martin. 2001. Word classes/parts of speech. In *International Encyclopedia of Social and Behavioral Sciences*. Edited by Paul B. Baltes and Neil J. Smelser. Amsterdam: Pergamon, pp. 16538–45.
- Heine, Bernd, Gunther Kaltenböck, Tania Kuteva, and Haiping Long. 2021. On the rise of discourse markers. In *Studies at the Grammar-Discourse Interface: Discourse Markers and Discourse-Related Grammatical Phenomena*. Edited by Alexander Haselow and Sylvie Hancil. Amsterdam: John Benjamins Publishing Company, pp. 23–55.
- Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51: 1205–47. [CrossRef]
- Horno Chéliz, María del Carmen. 2002. Aspecto léxico y verbos de percepción: A propósito de *ver* y *mirar*. In *In memoriam Manuel Alvar. Archivo de Filología Aragonesa*. Edited by Rosa María Castañer and José María Enguita. Zaragoza: CSIC, pp. 555–75.
- Kaltenböck, Gunther, Bernd Heine, and Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35: 848–93. [CrossRef]
- Kaltenböck, Gunther. 2016. Élément de macro-syntaxe: Comment catégoriser une classe insaisissable. *Modèles Linguistiques* 74: 11–28. [CrossRef]
- Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua & Stile* 20: 303–18.
- Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In *New Reflections on Grammaticalization*. Edited by Ilse Wischer and Gabriele Diewald. Amsterdam and Philadelphia: J. Benjamins, pp. 1–18.
- López-Couso, María José. 2010. Subjectification and intersubjectification. In *Historical Pragmatics*. Edited by Andreas H. Jucker and Irma Taavitsainen. Amsterdam: De Gruyter Mouton, pp. 127–63.
- Montolió Durán, Estrella, and Virginia Unamuno. 2001. The discourse marker *a ver* (catalan, *a veure*) in teacher-student interaction. *Journal of Pragmatics* 33: 193–208. [CrossRef]
- Nuyts, Jan. 2012. Notion of (inter)subjectivity. *English Text Construction* 5: 53–76. [CrossRef]
- Olbertz, Hella. 2012. The place of exclamatives and miratives in grammar—A functional discourse grammar view. *Revista Lingüística /Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro* 8: 76–98. Available online: <http://www.letras.ufrj.br/poslinguistica/revistalinguistica> (accessed on 24 November 2022).
- Padilla Herrada, Soledad. 2021. *Intervenciones Reactivas y Creación de Marcadores Discursivos: Un Enfoque Macrosintáctico*. Thesis doctoral, Universidad de Sevilla, Seville, Spain.
- Simeonova, Vesela. 2015. On the semantics of mirativity. Paper presented at the 2015 Annual Conference of the Canadian Linguistic Association Ottawa, Ottawa, ON, Canada, June 2–6.
- Slobin, Dan, and Ayhan Aksu. 1982. Tense, aspect and modality in the use of the Turkish evidential. In *Tense-Aspect: Between Semantics & Pragmatics*. Edited by Paul J. Hopper. Amsterdam: John Benjamins, pp. 185–200.
- Stein, Dieter, and Susan Wright, eds. 1995. *Subjectivity and Subjectivisation*. Cambridge: Cambridge University Press.
- Tanghe, Sanne, and Marlies Jansengers. 2014. Marcadores del discurso derivados de los verbos de percepción: Un análisis comparativo entre el español y el italiano. *Revue Romane* 49: 1–31. [CrossRef]
- Traugott, Elizabeth C., and Graeme Trousdale. 2010. Gradience, gradualness and grammaticalization: How do they intersect? In *Gradience, Gradualness and Grammaticalization*. Edited by Elizabeth C. Traugott and Graeme Trousdale. Amsterdam: John Benjamins, pp. 19–44.
- Traugott, Elizabeth C., and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Traugott, Elizabeth C., and Richard B. Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth C. 2003. From subjectification to intersubjectification. In *Motives for Language Change*. Edited by Raymond Hickey. Cambridge: Cambridge University Press, pp. 124–39.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Gradualness of Grammaticalization and Abrupt Change Reconciled: Evidence from Microvariation in Romance

Sandra Paoli

Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford OX1 2JD, UK; sandra.paoli@ling-phil.ox.ac.uk

Abstract: Grammaticalization has long been understood as a process that takes place gradually, but within it, discrete and abrupt changes take place. This tension has been reconciled by claiming that the semblance of a gradual process is given by different parts of a construction undergoing changes at different points in time. Focusing on synchronic microvariation as gradience, this article discusses cases of clitic loss in four Romance varieties (Brazilian Portuguese, Raeto-Romance, some northeastern Italo-Romance varieties, and French), and identifies common patterns in the cells of the paradigms that are most vulnerable to the process of loss. Relating the grammatical and semantic properties of these cells to established typological hierarchies, the paper explores how general cognitive principles can account for the key properties of gradualness and gradience and, ultimately, language change.

Keywords: gradualness; gradience; pronominal system; pronoun loss; Romance; features; animacy; person; specificity; case; number; hierarchies

1. Introduction

The idea that grammaticalization is a gradual process has long been at the forefront of conversations about language change, especially within more functional perspectives (Lichtenberk 1991; Hopper and Traugott 2003; Lehmann 2004, to name but a few). More recently, the edited volume by Traugott and Trousdale (2010) brought under the spotlight the relation between gradualness, gradience and grammaticalization: recasting the discussion in the context of a broader understanding of diachrony and synchrony as ‘two perspectives on the same thing’ (Lehmann 2004, p. 2), gradualness is seen as a manifestation of diachronic change, while gradience is the corresponding effect on the synchronic axis.

Reviewing Aarts’ (2004, 2007) distinction between two types of gradience, subjective and intersective, conceived, in broad terms, as gradience within and between categories, respectively, the authors conclude that such a sharp division is unwarranted, and the two can, in fact, be subsumed under subjective gradience. In Traugott and Trousdale’s understanding, then, gradience corresponds to a cline within a category ranging from more to less typical members of that category. This is spelt out more explicitly in Rosemeyer (2014, p. 78), who refines the definition of gradience as referring ‘to the fact that a construction displays structured variation regarding the use of lexical elements, with a cline from less to more typical elements’. What is of interest here is the specification that synchronic variation is structured, a distinction in terms of (proto)typicality, that is, between more core and more peripheral members, and the nature of the cline with respect to which they are measured.

Gradualness, according to Traugott and Trousdale (2010, pp. 25–26), is not to be interpreted as ‘an issue of indeterminacy, vagueness, or undecidability’, nor as ‘drift [. . .] or monotonic incremental change’, nor ‘equated with slow progression over long periods of time’, nor as ‘imperceptible change’: ‘[g]radualness refers to the fact that most change involves (a series of) micro-changes’ (Traugott and Trousdale 2010, p. 23), micro-steps that are discrete and therefore abrupt. They further address the tension between the two conceptions of grammaticalization, as a gradual process on the one hand, and as consisting

Citation: Paoli, Sandra. 2024.

Gradualness of Grammaticalization and Abrupt Change Reconciled: Evidence from Microvariation in Romance. *Languages* 9: 138. <https://doi.org/10.3390/languages9040138>

Academic Editors: Renata Enghels, Marlies Jansengers and Tom Bossuyt

Received: 7 December 2023

Revised: 16 March 2024

Accepted: 3 April 2024

Published: 9 April 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of discrete changes, involving abrupt shifts, on the other, concluding that ‘because different parts of a construction may undergo changes at different points in time, the change to the construction as a whole may appear to be gradual’ (Traugott and Trousdale 2010, pp. 20–21).

Diachronic gradualness and synchronic gradience are deeply intertwined because, ultimately, they are both expressions of the effects of the micro-changes that characterise language change: ‘some of these micro-steps may give rise to gradient systems at any synchronic “slice” in the development of a particular language’ (Traugott and Trousdale 2010, p. 39). It has long been believed that diachronic changes are manifested in synchronic variation (among others, (Andersen 2001, p. 228), and since change progresses ‘in an orderly, grammatically conditioned step-by-step fashion’ (Andersen 2001, p. 226), its synchronic manifestation is systematic (cf. Rosemeyer’s ‘structured variation’ mentioned above). Synchronic variation, in turn, gives rise to ‘change that has a determinate direction’ (Andersen 2001, p. 225). This two-way relation between synchrony and diachrony, that is, synchronic variation as both the outcome of and the trigger for diachronic change, adds a further dimension to an investigation of synchronic data, allowing the linguist to see reflected in synchrony the unfolding of language change.

Saussure’s polarised conception of synchrony and diachrony has long been abandoned, and although Traugott and Trousdale’s volume may be one of the first to directly and explicitly address the relation between the two, we find that typological cross-linguistic generalisations, such as Moravcsik’s (1974) verb agreement hierarchy, were long before then thought of and recast in diachronic terms as an indication of the order in which, for example, verbs develop agreement markings with their arguments (Givón 1976, p. 156; reformulated more overtly in Givón 2017, p. 90).

Selecting as case studies a set of closely related languages and investigating the synchronic microvariation that they display in the expression of pronominal object clitics, this article sets out to explore the linguistic factors that can be identified as responsible for the uneven progress from one stage of development to the next both within and across languages. More broadly, it explores how the key properties of both gradience and gradualness, highlighted above, can be derived. Unlike genetically unrelated languages, linguistic systems that share the majority of their characteristics and minimally differ from one another in a limited and identifiable number of properties, offer the opportunity to observe the variation manifested in a given phenomenon or structure without the interference of ‘background noise’ (cf., for example, Poletto 2012). The patterns with which morphological defectiveness expresses itself across these closely related languages are related to the Referential Hierarchy (in its original formulation by Silverstein 1976) and the Accessibility Hierarchy (Keenan and Comrie 1977), which have been shown to affect the morpho-syntactic organisation of a variety of languages (e.g., to do with number distinctions, differential case marking of transitive [\pm animate] subjects, and hierarchical verb agreement, as discussed in Comrie 1981, and references therein).

Taking a textbook example of grammaticalization as a starting point, this article explores specific instances in which the progress from one stage of development to the next has been uneven across the category affected by the phenomenon, in order to try and model the gradualness of grammaticalization in a system that is, ultimately, based on discrete categories and distinctions.

The grammaticalization process that we focus on is the developmental trajectory of pronominal forms as portrayed in the following cline (as, for example, expressed in (Fuß 2005, p. 4):

- (1) independent pronoun \rightarrow weak pronoun \rightarrow clitic pronoun \rightarrow affixal (agglutinative) agreement marker \rightarrow fused agreement marker $\rightarrow \emptyset$

This developmental path represents a specific instantiation of a more general tendency which is the line at the very core of grammaticalization: the shift from lexical and semantically contentful to grammatical and semantically empty to, eventually, complete disappearance (cf. Hopper and Traugott’s (2003, p. 7) cline content item > grammatical word > clitic > inflectional affix (>zero)). Once the final stages at the right edge of the cline

have been reached, languages often reintroduce a content element to replace the one that has disappeared, suggesting a more circular or spiral-like path than a linear progression (as already observed in the early twentieth century by scholars such as (von der Gabelentz 1901, p. 256; Meillet 1921, p. 140).

Focusing on the segment of the cline that represents the progress from clitic pronoun to zero, the investigation of the chosen case studies (Raeto Romance, Brazilian Portuguese, colloquial French, and a group of neighbouring Italo-Romance varieties spoken in north-eastern Italy, which have all been discussed in Paoli 2014) reveals a gradual and uneven progression of pronominal weakening that results in some cells of the clitic paradigm being affected before others. The identification of recurring patterns across the set of case studies allows us to explore the broader direction behind the phenomenon, which comes out of potentially universal hierarchies.

After a brief consideration of the emergence of pronominal forms from Latin to Romance in Section 2, the article turns to the four case studies, discussed individually in Sections 3.1–3.4, and comparatively, in Section 4. Some reflections and concluding remarks follow in Section 5.

2. The Emergence of Pronominal Clitics: From Latin to Romance

Before delving into the four case studies of clitic loss, it is helpful to briefly consider an earlier stage along the cline in (1), namely, the segment that represents the shift from independent pronoun to clitic; that is, the development of pronominal clitics. Latin did not have a clitic series of pronouns, and the development of pronominal clitics is considered one of the distinguishing morphosyntactic features that differentiate the Romance languages from Latin. The process that led to the birth of these atonic pronouns seems to have progressed unevenly (for a more detailed discussion of this stage, see Paoli 2014, pp. 167–69), reflecting perhaps also the change that was to come, the process of clitic loss.

Pronouns in their traditional conception were limited, in Latin, to the first and second persons, along with the reflexive form for the third person (as discussed in Wanner 1987, p. 67). Even then, their use was highly restricted, because Latin was, fundamentally, a null object (as well as a null-subject) language (cf., among many others, Johnson 1991; Mulder 1991; Luraghi 1997): conditions related to both syntactic and discourse restrictions operated on licensing the omission of objects. For the third person, however, when the relevant discourse and syntactic conditions allowed for it, the referents were lexicalised by demonstratives instead. While not systematically argued or independently substantiated, there are indications that in Latin two distinct prosodic versions of the same pronoun existed: often, a single Latin pronominal form evolved into two separate elements in Romance¹. At the pragmatic level, too, the same form could be used contrastively or simply refer to an expected/active referent, subsuming the two distinct roles that would be unpacked into the Romance tonic and clitic pronouns, respectively.

Given these facts, it is reasonable to infer that the Latin pronominal forms exhibited certain traits that would eventually define the characteristics of clitics in Romance. Wanner (1987, p. 68) highlights that the ‘true’ pronouns (referring to the first, second and reflexive third persons) had the capacity to exhibit clitic-like behaviour from their earliest usage. The demonstratives that were used as a third-person pronoun, on the other hand, were not witnessed in their clitic modern form until much later (i.e., around 800AD, in the early medieval parody of the *Lex Salica*; Wanner 1987, p. 68).

It therefore seems plausible to claim (as in Paoli 2014) that the (proto-)clitic use of first- and second-person forms preceded the same development of the third person, suggesting a hierarchy of the emergence of pronominal clitics based on [Person] features (‘>’ is to be understood as diachronic precedence):

- (2) 1st and 2nd > 3rd

The gradualness of the weakening of the tonic pronouns to produce clitics is hence due to the process targeting different cells of the paradigm at different times. Although not the specific focus of this article, it is interesting to note that this process of weakening

reveals a hierarchy based on Person. The contrast between first and second person, on the one hand, and third on the other, which is manifested in a range of widely differing linguistic phenomena, is well-attested cross-linguistically: as discussed in Silverstein (1976), Siewierska (2004), and related sources, third-person pronouns typically set themselves apart from their first- and second-person counterparts, which inherently carry indexical qualities, and whose reference depends on the identity of the speaker and listener. Benveniste (1971, p. 217) argues against grouping the third person with the first and second, considering it a ‘non-person’; Forchheimer (1953, pp. 5–6) identifies several morphological patterns highlighting how the unmarked nature of the third person allows for its distinct treatment compared to the first and second; Erteschik-Shir (2007) differentiates between the two in terms of discourse status, in that first and second person are ‘stage topics’ or ‘permanently available topics’, i.e., pronouns whose referents are always available in the discourse, unlike third person. Be whatever it may, what we see is that some more general, perhaps even universal, properties related to both structure and discourse impact the progress of an item along a grammaticalization cline. We return to this in Section 4.

Once all the cells of the paradigm had been affected, a complete series of clitic pronouns arose; following (Benincà 1995), I assume that this had taken place by the Middle Ages, at which point all (or most of) the Romance languages had developed a complete pronominal clitic series. This consideration is the basis on which the rest of this article is built, that is, that those modern varieties that do not have (fully) complete clitic paradigms have undergone (or are undergoing) a process of clitic loss.

3. Clitic Loss: Case Studies from Romance

Although generally considered a typical property of Romance, clitic pronouns are not found in all modern Romance languages: partially or totally defective clitic paradigms feature in the Raeto-Romance varieties spoken in the Grisons region (Graubünden) of Switzerland (Haiman and Benincà 1992), in Brazilian Portuguese (Galves 2000; Lucchesi and Lobo 1996; among others), and in a group of neighbouring northeastern Italian varieties (Paoli 2009, 2014), and the ability to omit an object clitic is established in colloquial French (Lambrecht and Lemoine 1996, p. 297; Larjavaara 2000, p. 63). Let us explore them in turn.

3.1. Brazilian Portuguese

Brazilian Portuguese (BP henceforth) allows pronominal objects of transitive verbs to remain unexpressed in constructions in which, in other Romance languages, a pronoun would be needed (cf., Cyrino and Reich 2002):

- | | | | | | | | |
|-----|----|--|-----------------|--------------|---------------|-------------|---------------|
| (3) | a. | <i>Você</i> | <i>tem</i> | <i>que</i> | <i>lavar</i> | \emptyset | <i>antes</i> |
| | | you | you-have | that | wash | \emptyset | before |
| | | <i>de</i> | <i>pôr</i> | <i>o!</i> | | | |
| | | of | put | \emptyset | | | |
| | | ‘You must wash (it) before putting (it) in!’ [talking to someone who is about to put rice into a pan] (from Cyrino 2017, p. 276) | | | | | |
| | b. | <i>Os</i> | <i>bolinhos</i> | <i>não</i> | <i>estão</i> | <i>aqui</i> | <i>porque</i> |
| | | the | cupcakes | not | they-are | here | because |
| | | <i>a</i> | <i>Maria</i> | <i>comeu</i> | \emptyset . | | |
| | | the | Maria | ate | \emptyset | | |
| | | ‘The cupcakes are not here because Maria ate (them)’ (from Cyrino 2017, p. 277) | | | | | |

Typologically speaking, null-object languages allow the omission of an object when it can be recovered from the context, either because discourse-old or because physically present in the space of speaker and hearer, and hence accessible. At the same time, BP also has a defective clitic paradigm in that the third-person forms have disappeared. The accusative clitics *o* ‘he/it.ACC.MASC.SG’, *os* ‘they.ACC.MASC.PL’, *a* ‘she/it.ACC.FEM.SG’ and *as* ‘they.ACC.FEM.PL’ do not exist in the spoken language; they are only found in written and formal contexts and are only acquired by children in school (Galves 2000, p. 147). Accusative clitic forms exist, however, for first and second person:

- (4) a. *Me chocou profundamente.*
 I.ACC.CL= it-shocked deeply
 'It deeply shocked me (from Duarte 2012, p. 317)
- b. *Te vi no cinema ontem.*
 you.ACC.CL= I-saw in-the cinema yesterday
 'I saw you at the cinema yesterday' (from Duarte 2012, p. 317)

In those contexts in which a third-person clitic pronoun would be otherwise expected, two strategies are available to BP speakers: they can either resort to a nominative tonic pronoun (used with accusative function) or leave the object unexpressed (as in (3) above). The choice is subject to the properties of the referent that acts as an antecedent: [+animate] referents must be expressed by a tonic pronoun (5a), while [-animate] referents can be left unexpressed (5b) (cf., Duarte 1989).

- (5) a. *Amanhã, o meu filho viajará para São Paulo.*
 tomorrow the my son he-will-travel to São Paulo
*Eu deixei ele/*ø ir sozinho.*
 I will-let he.TON/ø go alone
 'My son is leaving for São Paulo tomorrow. I am letting him go on his own'. (from Lucchesi and Lobo 1996, p. 308)
- b. *Eu comprei o dicionário e emprestei*
 I I-bought the dictionary and I-lent
*ø/*ele ao João.*
 ø/*it.TON to-the João
 'I bought the dictionary and lent it to João'. (from Lucchesi and Lobo 1996, p. 308)

A null object must be at least one of [-animate] or [-specific]: in the following examples in which both antecedents are [+animate], while the definite object *o preso* 'the prisoner' cannot be null, the non-specific *presos* 'prisoners' indeed can.

- (6) a. *O policial insultou o preso antes*
 the policeman he-insulted the prisoner before
*de torturar ele/*ø.*
 of torture he.TON/*ø
 'The policeman insulted the prisoner before torturing him' (from Vasconcellos Lopes and Cyrino 2005, p. 345)
- b. *O policial insulta presos antes de*
 the policeman he-insults prisoners before of
torturar ø/?eles.
 torture ø/?they. TON
 'The policeman insulted prisoners before torturing them' (from Vasconcellos Lopes and Cyrino 2005, p. 345)

In the dative, a third person is preferably expressed by *para* followed by a tonic form: compare (7a) to its European Portuguese (EP) counterpart (7b), which uses the clitic *lhe*.²

- (7) a. *Vou perguntar para ela.* (BP)
 I-go ask to she.TON
- b. *Vou-lhe perguntar.* (EP)
 I-go= she.CL ask
 'I will ask her'. (from Azevedo 2005, p. 236)

The shift has occurred gradually over the past few centuries. In a study examining plays from the 16th to the 20th centuries, Cyrino (1997) identifies a significant decline in the use of a third-person direct object in clitic form: from 89% in the 16th century to 21% in the 20th century. Initially, the null object was linked to a proposition, that is, a non-referential antecedent in the third person. Subsequently, this evolved to null objects referring to a [-animate] element, specifically a third-person, inanimate referent. Notably, animacy has played a fundamental role in dictating the sequence in which different cells of the paradigm are affected by this process of loss.

Cyrino et al. (2000) note that a dual evolution is underway in BP, as it is concurrently transitioning into a non-null subject language while also becoming a null-object language. Although the authors do not explicitly connect these processes, it is evident that both are being driven by the same set of features, albeit manifesting in opposing directions. Items categorized as [+human] and [+specific] tend to be phonetically realized as subjects first, while expletives such as ‘there’, which are [-human] and [-specific], remain null. Conversely, [-human] and/or [-specific] objects consistently remain unexpressed, that is, null. The authors identify this with the so-called Referential Hierarchy (Silverstein 1976; Comrie 1981; among others); we take up the discussion of this in Section 4.

3.2. Raeto-Romance

A number of closely related Raeto-Romance (R-R henceforth) varieties have varying degrees of defective clitic paradigms. As already discussed by Benincà and Poletto (2005, p. 228), the varieties spoken in Brigels and Camischollas have no pronominal clitics at all³: the accusative object in all the following examples is expressed with a form that is clearly tonic because it occupies a post-verbal position (8a); the possibility of it being enclitic on the verb is excluded by the fact that it can be separated from it by a variety of phrases (8b and 8c)⁴.

- | | | | | | | |
|-----|----|--|--------------|---------------|---------------|-------------|
| (8) | a. | <i>Ša</i> | <i>ti</i> | <i>vol</i> | <i>el.</i> | |
| | | if | you | want | it. | ACC.TON |
| | | ‘If you want it.’ (accusative, AIS VI:i, 1110) | | | | |
| | b. | <i>Yu</i> | <i>amfla</i> | <i>netur</i> | <i>el.</i> | |
| | | I | find | nowhere | it. | ACC.TON |
| | | ‘I cannot find it anywhere.’ (accusative, AIS VIII:ii, 1597) | | | | |
| | c. | <i>I</i> | <i>an</i> | <i>caciau</i> | <i>giodor</i> | <i>el.</i> |
| | | They | have | chased | away | he. ACC.TON |
| | | ‘They have chased him away.’ (accusative, AIS VIII:ii, 1667) | | | | |

These R-R varieties also completely lack a dative series, as well as a partitive (cf., Italian *ne* ‘of it’) and a locative clitic (cf., Italian *ci* ‘there’), which have been replaced, respectively, by the preposition *ad* ‘to’ followed by a tonic pronoun (9a), the preposition *kun* ‘with’ followed by a tonic pronoun (9b), and a null object (9c):

- | | | | | | | |
|-----|----|--|----------------|-------------|---------------|---------|
| (9) | a. | <i>Gi</i> | <i>kuai</i> | <i>ad</i> | <i>el.</i> | |
| | | say | that | to | he. | ACC.TON |
| | | ‘Tell him that.’ (dative, AIS VI:i, 1110) | | | | |
| | b. | <i>Koy</i> | <i>figesas</i> | <i>cun</i> | <i>el?</i> | |
| | | what | would-you-do | with | it. | ACC.TON |
| | | ‘What would you do with it?’ (partitive, AIS VIII, 1638) | | | | |
| | c. | <i>K</i> | <i>eu</i> | \emptyset | <i>mondi.</i> | |
| | | that | I | \emptyset | go | |
| | | ‘That I go (there).’ (locative, AIS VII, 1345) | | | | |

There are also R-R varieties that present a defective clitic paradigm, having retained an accusative series (10a, cf. the pre-verbal position), but completely lost the dative (10b), partitive (10c) and locative ones (10d). The process of clitic loss is still underway, and its gradual progression can be observed synchronically. An example of this is the variety spoken in Zernez:

- (10) a. *Ža* *tii* *il* *vovšt.*
 if you it. ACC.CL=
 'If you want it.' (accusative, AIS VI:i, 1110)
- b. *Di* *at* *el.*
 say to he. ACC.TON
 'Tell him.' (dative, AIS VIII:ii, 1659)
- c. *Ce* *fesat* *vus* *kun* *el?*
 what would-you-do you with it. ACC.TON
 'What would you do with it?' (partitive, AIS VI:i, 1113)
- d. *Voleys* *ka* *yow* \emptyset *geya?*
 you-want that I \emptyset go
 'Do you want that I go (there)?' (locative, AIS VIII:ii, 1638)

Based on this microvariation in the distributional patterns of accusative, dative, partitive and locative forms across this group of varieties, Benincà and Poletto (2005, pp. 227–28) draw a number of synchronic generalizations, and express them by means of an implicational hierarchy:

- (11) a. If a Romance language has clitics, it has direct object clitics.
 b. If a Romance language has dative clitics, it has direct object clitics.
 c. If a Romance language has partitive or locative clitics, it has dative clitics.

Given the premises laid out in Section 2, these synchronic generalizations can be interpreted as the reflection of a process of diachronic change (as suggested, for example, by Andersen 2001, p. 228), as indeed was done by Paoli (2014): the gradualness of grammaticalization is reflected in the synchronic microvariation across Case exhibited by the realisation of these pronominal forms in this group of closely related varieties. We can conclude that, here, the gradualness of grammaticalization is determined by Case: the partitive and locative clitics are affected first by the process of clitic loss, followed by the dative, followed, in turn, by the accusative paradigm. The hierarchy at play is therefore the following ('>' indicates diachronic precedence):

- (12) partitive/locative > dative > accusative

Some diachronic evidence adds some further finer distinctions to this Case hierarchy. A collection of texts dating from the 17th century (with some 18th-century amendments (Decurtins 1880–1883, p. 255)) from the R-R variety Surselvan, which, in its modern form, completely lacks a pronominal clitic series and uses instead the reintroduced tonic series, documents the presence of pronominal clitics. This suggests that, indeed, the current situation is the result of a gradual process of loss that has taken place over the last several hundred years and has followed a specific path. A close investigation of these texts, as in Paoli (2014, p. 172), confirms evidence of a Case hierarchy at work; in addition to the dative series having more cells that can be realised as null than the accusative series, it also offers evidence of the subsequent step, in which the original tonic forms are being reintroduced (recall that the cline representing the progression of pronominal loss goes from the clitic form, to zero, to the possible reintroduction of a tonic form). It also shows a further differentiation stemming from Person and animacy distinctions. Taking the third person as a point of reference, while the accusative paradigm can have a null object [-animate] referent in the plural, in the dative series, neither clitic nor null objects are found, and we only see the reintroduction of a tonic form, which represents a more advanced stage along the cline of grammaticalization in (1).

Furthermore, the same texts also show evidence of a further hierarchy based on Person and animacy features. As discussed in Paoli (2014, pp. 169–76), gaps in the paradigm are mostly found with third-person singular [-animate] referents, suggesting that this is the first form to be lost. These are followed by third-person singular [+animate], realised by tonic forms, in turn, followed by third-person plural. The process has not affected first- and second-person singular to the same extent; these are mainly realised by clitic forms, with a few null attestations. The scarcity of data prevents us from drawing any definite conclusions for first- and second-person plural. If considering Case, we see that in the dative paradigm, more cells can be realised as null compared to the accusative paradigm.

Considering the evidence from these 17th-century texts, we obtain the following partial hierarchies of loss for R-R based on Case (13a), Person and animacy features (13b), (>’ means, as above, diachronic precedence of loss):

- (13) a. dative > accusative
- b. 3rd singular [-anim > +anim] > 3rd plural [-anim > +anim] > 1st, 2nd singular

A final, additional reflection comes from the rather fluid situation witnessed in these 17th-century texts, in which clitic form, tonic form and null realisation can be used interchangeably. In (14), from Decurtins (1880–1883, p. 274, lines 5–6), the speaker refers to themselves by means of a clitic form first, followed in turn by a null object (under coordination reduction) and a tonic form. The pragmatics do not warrant the use of a tonic form (i.e., the tonic form *mei* does not express a contrast nor does it introduce a new referent), suggesting that all three realisations seem to be in free variation.

- (14) *Ti mi has giù ed ø has cha*
 you I.CL= you-have had and ø you-have not
saviu tener; ed ussa enqueres ti da
 been.able.to keep and now you-try you to
pigiari mei.
 keep I.ACC.TON
 ‘You had me and were not able to keep [me]; and now you are trying to detain me’.

The pronominal usage witnessed in these texts represents a snapshot of the change (that was still in progress in the 17th century) that Surselvan underwent. In addition to hierarchies based on both Case, Person and animacy, we see a situation of variation not dissimilar to the intermediate stage in cases of semantic change, in which both old and new meanings coexist; the interchangeability of the three realisations suggests that, at this stage, there are no firm boundaries between them, and although they remain morphologically (and possibly syntactically) distinct, their pragmatic distinctive features are blurred. Let us now turn to the northeastern Italo-Romance varieties, in which the interaction of Case, Person, Number and animacy paints a similarly complex picture.

3.3. Northeastern Italo-Romance

In a group of neighbouring Italo-Romance varieties spoken in the Dolomitic area in North-East Italy (described and analysed in Paoli 2009, 2014), the interaction of Case, Person, Number and animacy results in defective paradigms in both the accusative and dative clitic series.

In the accusative paradigm, there are clitic forms for first- and second-person singular and second-person plural⁵:

- (15) a. A: *Me sientes-tu?*
 I.ACC.CL= you-hear=SCL
 ‘Can you hear me?’
 B: *Si, te sient benon!*
 yes you.ACC.CL= I-hear well
 ‘Yes, I can hear you loud and clear!’ (from De Lorenzo Tobolo 1977, p. 347)
- b. *Tasè che duci ve sienti!*
 shut-up that all you.ACC.CL= they-hear
 ‘Shut up, everybody can hear you!’ (from De Lorenzo Tobolo 1977, p. 362)

For the third person, there are three possibilities, depending on whether the antecedent is animate or not: a clitic pronoun or, marginally, a tonic pronoun, for a [+animate] referent, or no pronoun at all if the referent is [-animate]. Although a [+animate] referent can be left unexpressed, neither a clitic nor a tonic pronoun can be used with a [-animate] referent:

- (16) a. A: *As-t* *vist* *Rosa?*
 have-you-SCL seen R
 ‘Have you seen Rosa?’
 B1: *Si,* *l’ei* *vista* *ngeri.*
 yes she.ACC.CL=I-have seen yesterday
 B2: *Si,* *ø ei* *vistu*⁶ *ngeri.*
 yes ø I-have seen yesterday
 ‘Yes, I saw her yesterday’ (from Paoli 2014, pp. 148–49)
 b. *A Maria* *piasi* *la pasta* *e ø mangia* *ogni di.*
 to M it-pleases the pasta and ø she-eats every day
 b'. **A Maria* *piasi* *la pasta* *e la* *mangia*
 to M it-pleases the pasta and she.ACC.CL= she-eats
 ogni di.
 every day
 ‘Maria likes pasta and she eats (it) every day’ (from Paoli 2014, p. 149)

A similar situation is found in the dative series, albeit with a higher level of deficiency in the paradigm; the only clitic forms that are still found and in use are for the second person, both singular and plural, and, marginally, for the first-person singular. For all the other cells, either a tonic form preceded by the preposition *a* ‘to’ or a null pronoun is used, the former being the most frequent choice. A summary overview of the accusative and dative clitic series is provided in Table 1:

Table 1. Accusative and Dative clitic series.

| | Accusative | Dative |
|--------|--|-------------|
| 1st sg | <i>me</i> | <i>me/ø</i> |
| 2nd sg | <i>te</i> | <i>te</i> |
| 3rd sg | <i>la/lu</i> [+anim]; <i>ø</i> [±anim] | <i>ø</i> |
| 1st pl | <i>ø</i> | <i>ø</i> |
| 2nd pl | <i>ve</i> | <i>ve</i> |
| 3rd pl | <i>?li</i> [+anim]; <i>ø</i> [-anim] | <i>ø</i> |

Against the premises laid out in Section 2, the incomplete clitic paradigms, with the dative series being more deficient than the accusative one, suggest that a process of clitic loss is ongoing, and that it is not uniformly targeting accusative and dative paradigms; once again, we see that there are some hierarchies at work, based on Case and Person. As for Case, the more numerous gaps in the dative paradigm suggest that the dative series of clitics has been affected first; the process of loss extended only subsequently to the accusative paradigm, producing a synchronic asymmetry that is very similar to the situation observed by Benincà and Poletto (2005) for the R-R varieties.

Drawing on the very restricted and, admittedly, not completely reliable diachronic evidence that exists for these northeastern Italo-Romance varieties, there is support for the interpretation of the synchronic incomplete paradigms as the result of a process of clitic loss. Despite these local varieties having been officially introduced in the drawing up of legal documents as early as 1631, instances of clitic pronouns in these texts cannot be found, as the legal writing style favours the repetition of the whole noun phrase rather than its replacement with a pronominal form. There is, however, a religious text translated possibly from Italian (or even Latin) that dates from around 1870, the Passion according to Saint Matthew (Tagliavini 1932), that offers a view of some atonic form usage. The text displays clitic forms for third-person singular and plural in the accusative paradigm, for both [+animate], (17a), and [-animate], (17b), referents; third-person singular dative clitics are used occasionally (but inconsistently, cf. 17b, 17c), suggesting a fluid situation in which the form was probably falling out of use and was being replaced by a null realisation. The third-person plural dative is only found as a null realisation (17c), while for the accusative,

there is a clitic (17d). The first-person plural dative is realised as a tonic pronoun preceded by a preposition (17e).

- (17) a. *Allora duggi i discepui i lu ha*
 then all the disciples SCL he.ACC.CL= they-have
impianteu vilò
 abandoned there
 'Then, the Disciples abandoned him there.' (Ch. XXVI, 56)
- b. *E Gesù el gh' rispondi: T*
 and Jesus SCL he.DAT.CL= he-answers SCL
l' has dit tu.
 it.ACC.CL= you-have said you
 'And Jesus replied: it is you who says it.' (Ch. XXVI, 64)
- c. *Giuda Scariott l' è du a ciattà i Capi*
 Judas Iscariot SCL he-is gone to find the chiefs
dei pré, e l' ø ha dittu ...
 of-the priests and SCL ø he-has said...
 'Judas Iscariot went to find the High Priests and said [to them] ...' (Ch. XXVI, 14)
- d. *I se ha divisu li me massari,*
 SCL reflex. they-have divided the my clothes
e i li ha tiredi alla sorti
 and SCL they.acc.cl= they-have thrown to-the fate
 'They shared my garments and cast lots for them.' (Ch. XXVII, 35)
- e. *Giuda l' ha purteu ndoi i trenta sodi*
 Judas SCL he-has taken back the thirty coins
ai Prenzipi dei Sazerdoti ma lueri
 to-the Princes of-the Priests but they
i ø ha rispondù Che importa
 SCL ø they-have answered what matters
a nujetar?
 to we.TON
 'Judas returned the thirty pieces of silver to the High Priests, but they answered [to him]:
 What does it matter to us?' (Ch. XXVII, 3–5)

Although the evidence afforded by these texts is not definitive, compounded with the synchronic facts, it forms a coherent picture; the process of clitic loss is ongoing, and, once again, we see that there are some hierarchies at work, based on Case, Person and animacy. As for Case, both synchronic and diachronic data indicate that the dative series of clitics was affected first; the process of decay extended only subsequently to the accusative paradigm, yielding a synchronic pattern that is very similar to what reported by Benincà and Poletto (2005) for the R-R varieties. Furthermore, we see a Person hierarchy at play, since first-person plural, followed by third person, are the cells that are most open to clitic loss. The complete absence of the clitic corresponding to the first-person plural suggests that this is the form that was targeted first by the process of clitic loss; this is, from a cross-linguistic perspective, an unusual fact (see Farrell 1990, pp. 329ff), as typically a null-object pronoun has a default interpretation of third person. Interpreting this fact from the perspective of a process of loss means that it is usually the third-person pronoun that is the first one to disappear. In view of this generalization, Paoli (2009, 2014) considers the first-person plural to be an outlier and discusses possible reasons why the first-person plural cell should be the one to show the highest vulnerability (Paoli 2009, pp. 78–79; 2014, pp. 165–66). The tentative explanation that is put forward there is based on the incompatibility between the semantic load of the composite form that is found in these varieties, a reflex of Latin NOS + ALTER (an exclusive 'we', which includes the speaker but excludes the hearer) and the weakness of a clitic form. Although no firm explanations can be reached, partly because of the scarcity of relevant evidence, a plausible conclusion that can be drawn is that whatever is happening to the first-person plural is not the result of the same process that is affecting the other cells of the paradigm. By removing the first-person plural pronouns from the

equation, the picture that emerges is not only more familiar, but also more congruous and meaningful: the process of decay affects the third person first, and within it, [-animate] referents are more readily expressed by a null form.

Summing up, the defective clitic paradigms in these northeastern Italo-Romance varieties reveal that two separate hierarchies determine the order in which cells are affected: the first one, based on Case, is represented in (18a); the second, based on Person, in (18b). Diachronic precedence of loss is, once again, expressed as '>':

- (18) a. dative > accusative
- b. 3rd sg [-anim > +anim] > 3rd pl [-anim > +anim] > 1st, 2nd sg > 2nd pl

Let us now turn to the final case study, colloquial French.

3.4. Colloquial French

Although French is not typologically classified as a null-object language, corpus-based research carried out in the last forty years or so (cf., Fónagy 1985; Lambrecht and Lemoine 2005; Larjavaara 2000, among others) has revealed systematic usages in which referential direct objects are not lexically realised:

- (19) a. A: *J'ai un truc pour toi si ça t'intéresse*
 I-have a thing for you if this interests you
 'I have something for you if you are interested'
- B: *C'est quoi?*
 it-is what
 'What is it?'
- A: *Je crois que t'aimes bien, toi,*
 I-think that you-love well you
 ce genre de truc. J'ai ø trouvé hier.
 this type of thing I-have ø found yesterday
 'I think that you like this sort of thing. I found (it) yesterday' (from Lambrecht and Lemoine 2005, p. 39)
- b. *Un jour, je me disais, je mettrais une petite annonce dans*
 one day I-told myself I-will-put a small ad in
 Le Provençal: [...]Mais je ø renvoyais toujours à plus tard.
 LP but I ø I-put-off always until later
 'One day, I told myself that I would put a classified ad in Le Provençal [...] But I kept putting (it) off until later'. (from Larjavaara 2000, p. 63)

Despite facing significant social stigma, this construction remains entirely viable in both casual written and spoken French. Its usage is governed by pragmatic, discursive and stylistic considerations, as discussed in works such as Cummins and Roberge (2005) and Lambrecht and Lemoine (2005). In their investigation of null objects in French, Cummins and Roberge (2005, p. 52) observe that the null object in instances such as those in (19) in which there is no overt clitic is always definite and referential; Lambrecht and Lemoine (2005, p. 14) note that the 'null instantiation of object pronouns is denoting specific discourse entities'. Null objects are, therefore, always definite, referential, specific, discourse-accessible and recoverable from the context; that is, they must be 'both active in the minds of the interlocutors at the time of utterance and [...] must have the status of an established discourse topic' (Lambrecht and Lemoine 2005, p. 42).

In addition, the null realization of an object is also dependent on further features of the referent, that is, Person and animacy: only third-person [-human] referents can be realised as null objects.

An interesting parallel emerges when considering the minimal pair in (20):

- (20) a. J' \emptyset *aime*.
 I \emptyset I-love
 'I love (it)'.
- b. Je l' *aime*.
 I s/he.ACC.CL= I-love
 'I love him/her'.

In any given context, (20a) can only refer to an inanimate referent that is discourse-accessible and recoverable from the context (for example, referring to a dress that a friend is trying on and showing the speaker, asking them for their views), while the object clitic l' in (20b) can only be used to refer to an animate antecedent (as in a conversation that is asking the speaker's opinion on a common friend).

The picture that emerges is, by now, very familiar; similarly to what we have already seen for Brazilian Portuguese and the northeastern Italo-Romance varieties, in colloquial French, too, it is third-person, [-animate] referents that can be left unexpressed. Assuming the very initial stages of a possible process of clitic loss, it can be seen clearly that the development is driven by a hierarchy based on animacy and probably Person.

4. Comparative Discussion

The investigation of the course of clitic loss across Brazilian Portuguese, northeastern Italo-Romance and colloquial French has brought to light strikingly similar patterns. Third-person cells are those that repeatedly appear to be the most vulnerable to clitic weakening and loss, with [-animate] referents preceding [+animate] ones in the process. The dative series, too, as witnessed in R-R and northeastern Italo-Romance, is affected by loss before the accusative series. Of the two trends, the hierarchy based on Person and animacy is the common denominator across all the varieties investigated. The hierarchy based on Case, on the other hand, is not immediately apparent; considering French as a manifestation of the incipient stage of the process of loss, we can speculate that a Case hierarchy may linearly follow (in terms of time) a Person and animacy one, or perhaps that they are to be interpreted as individual strands of a complex rope in which a number of hierarchies are simultaneously at play (much on the same line, for example, as Givón 1976, p. 52, or 2017, p. 91, who identifies a number of semantic hierarchies that are subsumed under a grammatical one based on grammatical function, closely reminiscent of Keenan and Comrie 1977 Noun Phrase Accessibility Hierarchy); the choice between these two possibly being determined by other properties of the language.

It was mentioned earlier that Cyrino et al. (2000) identify two unrelated changes that are concomitantly taking place in BP, a shift towards becoming a non-null subject and a null-object language, and convincingly make a case for both processes operating along the Referential Hierarchy (Silverstein 1976). Reflecting on the animacy hierarchy, Comrie (1981, p. 192) identifies the link between the natural world, grammatical features and related discourse properties: 'the animacy hierarchy [...] reflects a natural human interaction among several parameters, which include animacy in the strict sense, but also definiteness [...] and various means of making an entity more individuated—such as giving it a name of its own, and thereby making it more likely as a topic of conversation', that is, more accessible. Keenan and Comrie's (1977) Noun Phrase Accessibility Hierarchy, although formulated to capture the limitations on the availability of a noun phrase to being relativised, offers a ranking of grammatical roles that finds some correspondence in many other linguistic phenomena. As already argued by Paoli (2014, p. 179), these types of hierarchies, referential ones in particular, are an effective way to outline limitations on the morpho-syntactic behaviour of arguments. The limitations observed across languages regarding the positioning of noun phrases, and clitic and tonic pronouns emphasize the significance of referentiality in pronoun usage, whether expressed overtly or left null in languages that allow such possibilities. Similarly to what has already been noted by Comrie and Givón, referentiality emerges from a combination of factors: the nature of the nominal phrase, its argument status, Case, Person, animacy, ability to reference real-world entities,

and specificity. These features interact to establish a scale against which linguistic elements are assessed, creating, specifically, what is known as the Referential Hierarchy. At one end of the scale cluster are non-specific, non-referential, inanimate elements; at the other, specific, referential, human ones. Due to their inherent human and definite nature, first- and second-person pronouns rank higher than third person and are deemed ‘marked’. Among third-person pronouns, those categorized as [-human] naturally hold a lower rank than those marked as [+human], representing the most ‘unmarked’ forms. It seems to make intuitive sense that those linguistic forms that are least marked should be more accessible to a process of loss. Conversely, those forms that are richer in this respect, would be more resilient and disappear at a later stage (if at all). Some evidence that can lend support to this reasoning and the cognitive relevance of hierarchies in general, can be derived from preliminary neurolinguistic studies that suggest that referential hierarchies may hold the key to universal aspects of human cognition by playing an important role in language processing (e.g., Philipp et al. 2008; Wang et al. 2009). It is these general cognitive operations that, we claim, are ultimately responsible and determine the degree of affectedness, and, consequently, the order in which different parts of a construction undergo a given change.

In the Introduction, the notions of gradience and gradualness were discussed, and some keynote features were highlighted. For gradience, these were the structured nature of synchronic variation (Rosemeyer 2014, p. 78), its systematicity (Andersen 2001, p. 26), and the determinate direction of the change that it feeds into (Andersen 2001, p. 225). As for gradualness, it was defined as change that involves a series of micro-changes (Traugott and Trousdale 2010, p. 23) that progress in an orderly and grammatically conditioned fashion (Andersen 2001, p. 226). The link between these two notions explains how gradual change, which operates in a step-by-step manner, is manifested in synchronic variation that is systematic and structured. With respect to gradience, the concept of ‘typicality’, that is, a distinction between more core and more peripheral members of a given category, was also underlined; considering the instances of clitic loss under discussion, and the observations about first and second person, on the one hand, and third on the other, in Benveniste (1971, p. 217), Forchheimer (1953, pp. 5–6) and Erteschik-Shir (2007), we would conclude that while first- and second-person pronouns are more typical exponents of the category pronoun, third would be more peripheral. This distinction would then be relevant to the fact that third-person pronouns are more vulnerable to the process of loss; the least marked, or less typical, members of the category would be lost before the more marked, or more typical, ones. The understanding that general principles encoded in the Referential Hierarchy and the Noun Phrase Accessibility Hierarchy are the drivers of change, providing clear step-by-step paths of direction of change, explains the structure and systematicity of the synchronic manifestation of that change, and possibly even why less typical members of a category are more vulnerable to a process of loss.

In terms of gradience as the representation of microvariation within and between languages, the choice of genetically related languages has enabled a finer-grained observation of the modalities of action of the different grammatical and semantic properties that make up nominal elements such as Person, animacy, Case and specificity. The absence of so-called ‘background noise’ may have revealed an order of applicability, so to speak, that affects the way Case interacts with the other features with more clarity than would have been possible by adopting a classical typological approach, comparing genetically distant languages. The recurring patterns in the uneven realisation and spreading of the process of clitic loss unearthed in this article highlight the power of general (possibly even universal) principles, expressed in terms of both syntactic and semantic features whose operation can be captured in the form of an ordered scale. Irrespective of how exactly hierarchies are formulated and labelled, by representing properties that transcend the specific, both in terms of language-specific and phenomenon-specific, they can explain morphosyntactic patterns and the way they are actualised across space and time.

5. Conclusions

Starting from Traugott and Trousdale's explicit connection between gradualness of change and synchronic gradience, this study set off to explore what may lie behind cases of uneven change. Taking as its object of research a series of cases of clitic loss in a set of Romance varieties, the investigation has brought to light a common denominator, that is, the role played by Person, animacy, specificity and Case in the process of loss. All these are properties that find an expression in the Referential Hierarchy and the Noun Phrase Accessibility Hierarchy, which may be possibly collapsed into the grammatical function hierarchy subject > direct object > oblique (Givón 2017, p. 91). In addition to being responsible for the process of loss itself, as already claimed by Paoli (2014), this article has made a case for the same general aspects of language to be also responsible for the gradualness of change, its determinate direction, and an explanation for the systematicity and structure of its synchronic manifestation. This is particularly evident when comparing diachronic change in closely related languages.

Although the link between the synchronic deficiency of the clitic paradigms of the Romance varieties investigated here and the gradual process of loss is quite clear in this particular case, Traugott and Trousdale conclude that not all instances of synchronic gradience are the result of grammaticalization (Traugott and Trousdale 2010, p. 39). Further research on a wider range of phenomena in a larger number of languages (both typologically distinct and genetically related) is needed in order to explore criteria that may be used to establish whether synchronic gradience is the result of grammaticalization or not. There is also a need for a deeper understanding of the nature and properties of hierarchies in general. However, it seems that the application of hierarchies and, by reflex, of general cognitive principles, may be a promising way to understand the essence of language change.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data discussed in this article are available in the relevant published papers or in the Linguistic Atlas cited within.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ In Old French, for example, there were two first-person singular forms in the accusative, the tonic *mei/moi* and the atonic *me*: these, a clear development from Latin *MĒ* 'I.ACC', likely arose from stressed and unstressed usage, respectively, resulting in the anticipated transformation of tonic [e] into the diphthong ei/oi and atonic [e] into a [ə] (Salvi 2001, pp. 286ff.).
- ² *Lhe* has now come to function in spoken BP as a second-person clitic, both for the dative and the accusative.
- ³ The data in this section were extracted from the *Atlante Italo-Svizzero/Sprach und Sachatlas Italiens und der Südschweiz* 'Italo-Swiss Atlas' (AIS) (Jaberg and Jud 1928–1940): the numbers indicate the volume (in Roman numerals) followed by the map on which that specific stimulus is found. The AIS data were collected through eliciting the translation of a list of sentences from Italian into the varieties spoken in each of the locations: in the original Italian sentence, the clitics are present, and the R-R speakers are notably omitting them in spite of the stimulus.
- ⁴ The information given in the glosses is simplified, and focuses on the type of pronoun used. Hence, only the following abbreviations are used: CL 'clitic/atonic pronoun', TON 'tonic pronoun', Ø 'no clitic used/null realisation', SCL 'subject clitic', ACC 'accusative', DAT 'dative'. For simplicity the 'Ø' symbol is placed where a clitic form would be expected; this seems to be reasonable given that all the varieties considered here have tonic forms, and a non-realized form is certain to be a (missing) clitic.
- ⁵ For the first-person plural there exists no clitic form: a first-person plural antecedent can either be referred back to with the tonic form or with no pronoun at all. Considering this pattern within the broader picture of the other cases of clitic loss, Paoli (2009, 2014) assumes the first-person plural to be an outlier.
- ⁶ In (16a B1), the past participle *vista* 'seen' agrees in number and gender with its pre-verbal accusative clitic. No such agreement appears when the clitic is omitted, as in (16a B2), suggesting that the form is not just phonetically unrealised but syntactically inactive, hence totally absent.

References

- Aarts, Bas. 2004. Modelling linguistic gradience. *Studies in Language* 28: 1–49. [CrossRef]
- Aarts, Bas. 2007. *Syntactic Gradience. The Nature of Grammatical Indeterminacy*. Oxford: Oxford University Press.
- Andersen, Henning. 2001. Actualization and the (uni)directionality. In *Actualization: Linguistic Change in Progress*. Edited by Henning Andersen. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 225–48.
- Azevedo, Milton. 2005. *Portuguese: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Benincà, Paola, and Cecilia Poletto. 2005. On some descriptive generalizations in Romance. In *Oxford Handbook of Comparative Syntax*. Edited by Guglielmo Cinque and Richard S. Kayne. Oxford: Oxford University Press, pp. 221–58.
- Benincà, Paola. 1995. Complement Clitics in Medieval Romance: The Tobler-Mussafia Law. In *Clause Structure and Language Change*. Edited by Adrian Battye and Ian Roberts. New York and Oxford: Oxford University Press, pp. 325–44.
- Benveniste, Emile. 1971. *Problems in General Linguistics*. Coral Gables: University of Miami Press.
- Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. Oxford: Blackwell.
- Cummins, Sarah, and Yves Roberge. 2005. A modular account of null objects in French. *Syntax* 8: 44–64. [CrossRef]
- Cyrino, Sonia. 1997. *O Objeto nulo no Português do Brasil—um Estudo Sintático-Diacrónico*. London: Editora da UEL.
- Cyrino, Sonia. 2017. On animacy restrictions for the null object in Brazilian Portuguese. In *Contrastive Studies in Verbal Valency*. Edited by Lars Hellan, Andrej L. Malchukov and Michela Cennamo. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 275–94.
- Cyrino, Sônia, and Uli Reich. 2002. Uma visão integrada do objeto nulo no Português Brasileiro. *Romanistisches Jahrbuch* 52: 360–94. [CrossRef]
- Cyrino, Sônia, Maria Eugênia Duarte, and Mary Kato. 2000. Visible subjects and invisible clitics in Brazilian Portuguese. In *Brazilian Portuguese and the Null Subject Parameter*. Edited by Mary Kato and Esmeralda Vailati Negrão. Frankfurt and Madrid: Vervuert-Iberoamericana, pp. 55–104.
- Decurtins, Caspar, ed. 1880–1883. Quattro testi soprasilvani. *Archivio Glottologico Italiano* 7: 151–364.
- De Lorenzo Tobolo, Elia. 1977. *Dizionario del dialetto ladino di Comelico Superiore*. Bologna: Edizioni Tamari.
- Duarte, M. Eugênia. 1989. Clítico acusativo, pronome lexical e categoria vazia no português do Brasil. In *Fotografias Sociolinguísticas*. Edited by Fernando Tarallo. Campinas: Editora da UNICAMP, pp. 19–34.
- Duarte, M. Eugênia. 2012. When speech and writing are too far apart. Non-dominant features of Brazilian Portuguese becoming dominant. In *Non-Dominant Varieties of Pluricentric Languages. Getting the Picture*. Edited by Rudolf Muhr. Frankfurt am Main: Peter Lang, pp. 315–26.
- Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. New York and Oxford: Oxford University Press.
- Farrell, Patrick. 1990. Null objects in Brazilian Portuguese. *Natural Language & Linguistic Theory* 8: 325–46.
- Forchheimer, Paul. 1953. *The Category of Person in Language*. Berlin: Walter de Gruyter.
- Fónagy, Ivan. 1985. J’aime, je connais: Verbes transitifs à objet latent. *Revue Romane* 20: 3–35.
- Fuß, Erik. 2005. *The Rise of Agreement, a Formal Approach to the Syntax and Grammaticalization of Verbal Inflection*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Galves, Charlotte. 2000. Agreement, predication and pronouns in the history of Portuguese. In *Portuguese Syntax: New Comparative Studies*. Edited by João Costa. New York and Oxford: Oxford University Press, pp. 143–68.
- Givón, Talmy. 1976. Topic, Pronoun and Grammatical Agreement. In *Subject and Topic*. Edited by Charles Li. New York: Academic Press, pp. 149–88.
- Givón, Talmy. 2017. *The Story of Zero*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Haiman, John, and Paola Benincà. 1992. *The Rhaeto Romance Languages*. London: Routledge.
- Hopper, Paul, and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Jaberg, Karl, and Jakob Jud. 1928–1940. *Atlante Italo-Svizzero/Sprach und Sachatlas Italiens und der Südschweiz*. Zofingen, Bern. Available online: <https://navigais-web.pd.istc.cnr.it> (accessed on 27 November 2023).
- Johnson, Ruth. 1991. The Direct Object Pronoun as Marker of Transitivity in Latin. Ph.D. dissertation, UCLA, Los Angeles, CA, USA.
- Keenan, Edward L., and Bernard Comrie. 1977. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8: 63–99.
- Lambrech, Knud, and Kevin Lemoine. 1996. Vers une grammaire des compléments d’objet zero en français parlé. In *Absence de marques et représentation de l’absence*. Edited by Jean Chuquet and Marc Fryd. Rennes: Presses Universitaires de Rennes, pp. 279–309.
- Lambrech, Knud, and Kevin Lemoine. 2005. Definite null objects in (spoken) French: A construction-grammar account. In *Grammatical Constructions: Back to the Roots*. Edited by Mirjam Fried and Hans C. Boas. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 13–55.
- Larjavaara, Meri. 2000. *Présence ou Absence de l’objet: Limites du Possible en Français Contemporain*. Helsinki: Academia Scientiarum Fennica. Available online: <https://helda.helsinki.fi/server/api/core/bitstreams/215f013c-eed4-48d2-b7ac-05e75c361fd2/content> (accessed on 27 November 2023).
- Lehmann, Christian. 2004. Theory and method in grammaticalization. *Zeitschrift für Germanistische Linguistik* 32: 152–87. [CrossRef]
- Lichtenberk, Frantisek. 1991. On the gradualness of grammaticalization. In *Approaches to Grammaticalization*. Edited by Elizabeth Closs Traugott and Bernd Heine. Amsterdam: John Benjamins Publishing Company, vol. 1, pp. 37–80.
- Lucchesi, Dante, and Tânia Lobo. 1996. Aspectos da sintaxe do Português Brasileiro. In *Introdução à Linguística geral e Portuguesa*. Edited by Isabel Hub Faria, Emília Ribeiro Pedro, Inês Duarte and Carols A. M. Gouveia. Lisboa: Caminho, pp. 303–22.

- Luraghi, Silvia. 1997. Omission of the Direct Object in Classical Latin. *Indogermanische Forschungen* 102: 239–57.
- Meillet, Antoine. 1921. L'évolution des formes grammaticales. In *Linguistique Historique et Linguistique Générale*. Paris: Honoré Champion, pp. 130–48.
- Moravcsik, Edith. 1974. *Object-Verb Agreement*. Working papers on language universals 15. Stanford: Stanford University, pp. 25–140.
- Mulder, Hotze. 1991. Object Continuity and Elliptical Objects in Latin. *Romance Linguistics and Literature Review* 4: 12–26.
- Paoli, Sandra. 2009. Gerarchie di Caso e Persona nei paradigmi dei pronomi atoni obliqui: Il caso dei dialetti del Comelico. *Rivista Italiana di Dialettologia* 33: 59–84.
- Paoli, Sandra. 2014. Defective object clitic paradigms and the relation between language development and loss. *Journal of Linguistics* 50: 143–83. [CrossRef]
- Philipp, Markus, Ina Bornkessel-Schlesewsky, Walter Bisang, and Matthias Schlewsky. 2008. The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain and Language* 105: 112–33. [CrossRef] [PubMed]
- Poletto, Cecilia. 2012. Contrastive linguistics and micro-variation: The role of dialectology. *Languages in Contrast* 12: 47–68. [CrossRef]
- Rosemeyer, Malte. 2014. *Auxiliary Selection in Spanish: Gradience, Gradualness, and Conservation*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Salvi, Giampaolo. 2001. La nascita dei clitici romanzi. *Romanische Forschungen* 113: 285–319.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical Categories in Australian Languages*. Edited by Robert M. W. Dixon. Canberra: Australian Institute of Aboriginal Studies, pp. 112–71.
- Tagliavini, Carlo. 1932. Testi dialettali comelicesi del secolo scorso. *Rivista Italiana di Letteratura Dialettale* IV: 109–23.
- Traugott, Elizabeth Closs, and Graeme Trousdale. 2010. *Gradience, Gradualness and Grammaticalization*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Vasconcellos Lopes, Ruth, and Sônia Cyrino. 2005. Animacy as a driving cue in change and acquisition in Brazilian Portuguese. In *Linguistic Evidence—Empirical, Theoretical and Computational Perspectives*. Edited by Stephan Kepser and Marga Reis. Berlin: Mouton de Gruyter, pp. 85–102.
- von der Gabelentz, Georg. 1901. *Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse*, 2nd ed. Leipzig: Chr. Herm. Tauchnitz.
- Wang, Luming, Matthias Schlewsky, Balthasar Bickel, and Ina Bornkessel-Schlewsky. 2009. Exploring the nature of the “subject”-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes* 27: 1180–226. [CrossRef]
- Wanner, Dieter. 1987. *The Development of Romance Clitic Pronouns: From Latin to Old Romance*. Berlin: Mouton de Gruyter.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Numeral Incorporation as Grammaticalization? A Corpus Study on German Sign Language (DGS)

Felicitas Otte *, Anke Müller, Sabrina Wähl and Gabriele Langer

Institute of German Sign Language and Communication of the Deaf, University of Hamburg, 20354 Hamburg, Germany; anke.mueller@uni-hamburg.de (A.M.); sabrina.waehl@uni-hamburg.de (S.W.); gabriele.langer@uni-hamburg.de (G.L.)

* Correspondence: felicitas.otte@uni-hamburg.de

Abstract: Numeral incorporation describes the merging of a numeral sign with a lexical sign to create a single sign with a compositional meaning, e.g., “three weeks.” As a phenomenon of simultaneous morphology, numeral incorporation is unique to sign languages. While researchers disagree on the exact morphological structure of the construction, it has, thus far, mainly been described as a synchronic, phonological phenomenon. Using the DGS corpus, a language resource on German Sign Language, we explore the possibility of numeral incorporation resulting from a language change process, specifically a grammaticalization process. Our dataset comprises tokens belonging to nine different signs that may occur in numeral incorporations. We find a cline of three constructions in the corpus, which shows a progression from free morpheme to cliticized morpheme to bound morpheme (affix). A comparison of the usage frequency of the three constructions in different age groups reveals that signers use more incorporations the younger they are. Following the apparent time approach, these observations are taken as indicators of diachronic language change. We describe to what extent the properties of numeral incorporation fit with the grammaticalization hypothesis and conclude that while the emergence of numeral incorporation is an instance of language change and shows some aspects seen in grammaticalization, the gradual change fails to exhibit some crucial aspects of grammaticalization and, thus, should not be regarded as an example thereof.

Keywords: German Sign Language; numeral incorporation; grammaticalization; corpus study

Citation: Otte, Felicitas, Anke Müller, Sabrina Wähl, and Gabriele Langer. 2023. Numeral Incorporation as Grammaticalization? A Corpus Study on German Sign Language (DGS). *Languages* 8: 153. <https://doi.org/10.3390/languages8020153>

Academic Editors: Renata Enghels, Marlies Jansengers and Tom Bossuyt

Received: 15 February 2023

Revised: 1 June 2023

Accepted: 4 June 2023

Published: 19 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Numeral incorporation is a very common phenomenon in sign languages and denotes the integration of a numeral sign into a lexical sign by merging features of both into a new sign form. In German Sign Language (DGS), for example, temporal signs such as WEEK1A “week” can incorporate handshape features of a restricted set of numerals to denote the meaning “quantity of unit,” e.g., “quantity of weeks.” In the past, numeral incorporation has been described as a synchronic, morpho-phonological phenomenon (Chinchor 1982; Mathur and Rathmann 2011; Semushina and Mayberry 2019). In this study, we want to explore the possibility of numeral incorporation being the result of diachronic language change, specifically, a grammaticalization process. Our study is based on the recently compiled DGS corpus, which contains videos of signed language data. As there is no suitable diachronic data available for DGS, we employ the synchronic data of the DGS corpus to compare the use of numeral incorporation across different age groups (using the apparent time approach). Following the usage-based framework of construction grammar, we can deal with heterogeneous data showing multiple and, at times, competing constructions at work, e.g., sequential and simultaneous constructions.

Our results confirm that numeral incorporation is a field with semantic and phonological constraints, but the data also indicates ongoing language change: frequency of use and phonological change may lead to an increase in use of numeral incorporation in terms of quantity, but also in terms of attracting more lexical signs to the pool of signs

allowing for numeral incorporation. For some signs, the language change process leading to numeral incorporation is incomplete and will need research on future data to confirm the entrenchment—or abandonment—of the construction.

2. Theoretical Background

In this section, we provide some background on the phenomena relevant to our study. We start with an introduction to the numeral systems in sign languages and DGS, in particular (Section 2.1), continue with a description of numeral incorporation (Section 2.2), and, lastly, provide some background information on grammaticalization (Section 2.3.).

2.1. Numeral Systems in Sign Languages

Before describing the specifics of the DGS numeral system, we provide a brief introduction to sign language phonology, which will be helpful in following the subsequent descriptions. Signs can be generally described as consisting of four different parameters: handshape, location, hand orientation, and movement. As Brentari (2012, p. 22) described it, parameters are “different phonemic groups of features.” Thus, they constitute contrastive properties in the form of a sign. Specifically, a parameter can only be specified for one (possibly complex) value at a time. In a fully specified sign, all parameters are specified in the phonemic representation and are realized in the articulation of a sign. The phonological status of these parameters is shown by minimal pairs, as illustrated in Appendix A, Figure A1. Changing the value of one parameter of an individual sign, e.g., the hand orientation of MILK2C “milk,” leads to a different sign, e.g., DIFFICULT1 “difficult.”

Sign languages differ in their numeral systems with respect to the underlying count base, as well as to the motivation of forms and productive strategies to form numerals higher than 10, and whether they employ one or two hands. Sagara and Zeshan (2016, pp. 28–33) give a typological overview of the cardinal number system of 30 sign languages. Numeral systems are often a mixture of lexical and productive word forms.

Most of the larger or urban sign languages have a base of 10, which is similar to the majority of spoken languages. This may be due to the number of fingers of the human hands (Zeshan et al. 2013, p. 360). Numeral signs can reflect different iconic motivations, e.g., form iconicity with respect to the written number (which Sagara and Zeshan (2016, p. 28) call orthographic iconicity) or to the initial letter of the written spoken language numeral, or different kinds of a finger-for-number analogy for the numbers 1 to 10. In a two-handed system, the second hand is added to enumerate 6–10, while in a one-handed system, the numbers six to nine or 10 can be articulated on one hand as well, e.g., by touching a certain finger with the thumb.

The numeral system of DGS is two-handed, and the numerals for the numbers 1–10 are “based on ‘number-for-number iconicity’ (Taub 2001), that is, the number of extended fingers corresponds to the numerical value to be expressed” (Sagara and Zeshan 2016, p. 29). The selection of fingers results in specific handshapes serving to distinguish the numerals from 1–10. These are called numeral handshapes. DGS numerals from 1 to 10 vary based on which fingers are extended (e.g., “one” can be expressed by either an extended index or an extended thumb) and the hand orientation (see Konrad et al. 2022, pp. 14–16, for more detail). Variants in hand orientation either have the palm directed away from or towards the body. Two-handed signs (from 6–10) may exhibit mixed orientations; that is, with one of the hands directed away from the body and the other directed towards the body. All numerals from 1–10 share one orientation feature: the hand is upright, and the extended fingers point upwards. For pictures of the numeral handshapes 1–10, see Appendix B, Figure A2, which include variants in handshape but leave out variance in palm orientation (palm directed away or towards the body).

Numerals for numbers higher than 10 are usually complex numerals; that is, they are formed by syntactic or morphosyntactic rules (von Mengden 2010, p. 28). In DGS, and in sign languages in general, there are two basic categories: numerals expressed in a single, morphologically complex sign, or in a sequence of signs. Numerals of the first group are

expressed in a regular way by combining a specific movement and hand orientation with selected extended fingers to express values from 1 to 9. Sagara and Zeshan (2016, p. 29) describe this strategy as “combining a numeral handshape with a numeral movement pattern.” The teens, i.e., the numbers from 11 to 19¹, are formed combining the numeral handshapes from 1 to 9, with a meaningful movement signifying the addition of 10 (see Appendix C Example (1), for an example featuring a TEEN sign). Multiples of 10 are likewise expressed, combining numeral handshapes from 2 to 9, with another movement signifying the multiplication with 10 to produce numerals from 20 to 90. The multiples of hundred and thousand are also expressed applying the same strategy. The numeral handshapes from one to nine/ten can be attached to a numeral movement to signify 100 up to 900, and 1000 up to 10,000. Figure 1 illustrates the numerals 100, 200, and 900. Tens, hundreds, and thousands differ with regard to their movements.

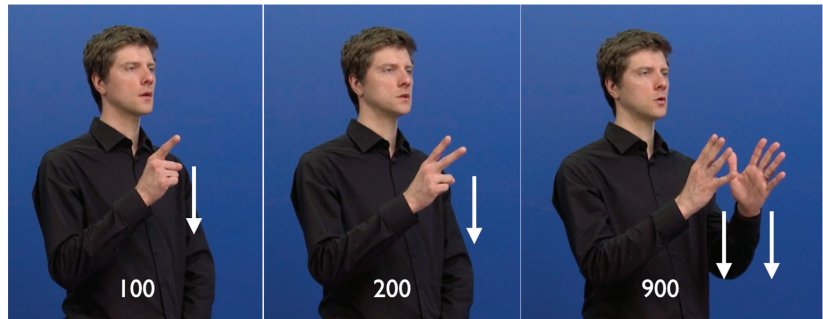


Figure 1. Productive strategy for multiples of 100.

Millions, on the other hand, are complex numerals that are expressed sequentially; that is a numeral from 1 to 10 followed by a sign for million, meaning the multiplication of millions by the preceding number. All numbers that express high values by combining lower values are expressed through a mixture of morphosyntactic strategies. They are not relevant for this paper but should be mentioned for completeness. A number like 47 is signed sequentially in DGS: first the numeral for seven, followed by the numeral for 40 (additive strategy). This order mirrors the practice in spoken German. Example (2) in Appendix C illustrates the use of such a numeral.

While the term complex numerals denotes all of the above described numerals higher than 10, from now on, we restrict the use of the term to complex numerals of the first kind, that is to complex DGS numerals with simultaneous morphology encompassing the teens, the multiples of ten, hundred, and thousand.

Another remark on terminology may be in order. The terms “numeral sign” and “numeral” refer to the same concept, the linguistic expression of a number. If we want to refer to the handshape parameter of numerals only, we speak of numeral handshapes as opposed to numeral signs or numerals (with all parameters specified).

2.2. Numeral Incorporation

Numeral incorporation is the simultaneous combination of a numeral sign and a lexical base sign. As a linguistic structure, it is common to many sign languages (Sagara and Zeshan 2016, pp. 31–32). Lexical signs that allow numeral incorporation usually belong to semantic fields such as time units, calendric terms, measurements, currencies, and school grades. However, the specific set of signs that do incorporate numerals depends on the language (Sagara and Zeshan 2016, p. 32; Ktejik 2013, p. 208; Mathur and Rathmann 2011, p. 62; Chinchor 1982, p. 77). Another restriction pertains to which numerals can be incorporated. This limitation seems to be of a phonological nature, depending on the phonological features of the numeral or the numeral system of a sign language (see Sagara and Zeshan 2016 for a typological overview). In DGS, the numerals 1 to 10

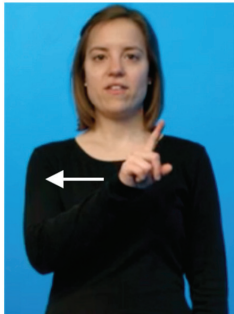
can be incorporated,² whereas Semushina and Mayberry (2019, pp. 109–10) report the incorporation of the numerals from 1 up to 15 in Russian Sign Language (RSL), at least for one sign. Phonological features of the incorporating lexical base sign play a role as well (Semushina and Mayberry 2019; Mathur and Rathmann 2011), e.g., restrictions on the handshape parameter. We will elaborate on this a little bit later in this section.

The following example in Figure 2a,b shows a sequential expression with the meaning of “three weeks.” This construction stands in contrast to the numeral incorporation with the same meaning shown in Figure 3. We will use this example to illustrate the morpho-phonological process behind numeral incorporation. As the exact nature of the process is still under discussion, we describe several perspectives.



| | |
|--------------------|---|
| Handshape | “3-handshape:” fist with extended thumb, index, and middle finger |
| Location | Centralized in front of the body |
| (palm) Orientation | Towards the body |
| Movement | Slight forward movement |

(a)



| | |
|--------------------|--|
| Handshape | “Index-handshape:” fist with extended index finger |
| Location | Centralized in front of the body |
| (palm) Orientation | Away from the body |
| Movement | Linear movement to the right (left when produced with the left hand) |

(b)

Figure 2. (a). Phonological parameters and illustration of the DGS sign THREE1A. (b). Phonological parameters and illustration of the DGS sign WEEK1A.



| | |
|--------------------|--|
| Handshape | “3-handshape:” fist with extended thumb, index, and middle finger |
| Location | Centralized in front of the body |
| (palm) Orientation | Away from the body |
| Movement | Linear movement to the right (left when produced with the left hand) |

Figure 3. Phonological parameters and illustration of the incorporated form “three weeks.”

Figure 2a,b show the fully specified (non-incorporated) signs signed in sequence (the numeral THREE1A³ “three” and the lexical sign WEEK1A “week”) and describe each sign’s phonological parameters.

When signed sequentially, both signs realize all of their phonological parameters; that is, they are fully specified signs with their own specific movements, locations, orientations, and handshapes. In a numeral incorporation, in contrast, there is only one sign in which the numeral and the lexical base sign are merged. As a phonological parameter can be realized only through one value at a time, some values in the incorporated form are determined by the numeral, while others are determined by the base sign and as a consequence; some of the individual signs’ features are deleted. The result is a morphologically complex form (Figure 3) that takes on the handshape of the numeral but retains the specifications of the base sign in the other three parameters⁴.

As for the analysis of numeral incorporation, various explanations have been proposed for different sign languages. In an early analysis of numeral incorporation in American Sign Language (ASL), Chinchor (1982) describes the sign formation process of numeral incorporation as the combination of two signs. These are not both fully specified signs, though, as the numeral sign is regarded as unspecified for location and movement (Chinchor 1982, p. 130). Chinchor, furthermore, claims that the base sign (in its unmodified non-incorporating form) has to have a handshape matching the numeral “one” in one of its allomorphs. This rule has also been stated by Mathur and Rathmann (2011) as a tendency in both ASL and DGS.

Mathur and Rathmann (2011) describe numeral incorporation as a morphological process where a numeral and a sign with the above-mentioned semantic properties become one sign. Both signs have “a fixed phonological realization” (Mathur and Rathmann 2011, p. 57) in their fully specified form. In the process of merging, the numeral lends its handshape and loses its features of placement and movement, while the lexical base sign loses its handshape parameter and lends its other parameter values to the resulting construction. Mathur and Rathmann (2011) place this process within the group of non-concatenative word formation, as the articulation of the relevant features is simultaneous rather than a sequence of morphemes. Which signs take part as base signs in the process could be specified in a list in the lexicon (comparable to lists of irregular verbs with vowel change as for instance the past tense morpheme in English). However, there are phonological reasons for why some signs incorporate numerals and others do not. Thus, it seems that if the lexical sign has a marked handshape⁵, the handshape cannot be deleted; if the numeral has a complex, distinctive movement feature, this movement likewise cannot be deleted. In two-handed signs, symmetry constraints may prevent incorporation (Mathur and Rathmann 2011, pp. 67–69). Semushina and Mayberry (2019) attest and refine these general findings for RSL.

A contrary morphological analysis is proposed by Liddell (1996), who describes incorporated forms as consisting of a bound root as a lexical base and a numeral morpheme. The bound root is specified for all phonological parameters, except for the handshape, while the numeral morpheme consists of only a handshape and, as such, is also a bound morpheme. The two morphemes are then combined to form one sign with four specified phonological parameters. This analysis is based on the fact that some incorporated forms do not have a non-incorporated, independent base sign with four phonological parameters, e.g., the ASL sign O’CLOCK (“o’clock”) is always specified for a specific time, i.e., it always contains a numeral handshape. Thus, it does not exist independently as a fully specified base sign. Aiming at finding a theory that accounts for these, as well as the signs with fully specified independent base signs, Liddell argues that incorporated forms as the one shown above for “week” are not based on the fully specified sign, but they are likewise formed with a bound root, which matches the fully specified sign in all parameters but a handshape (as it has an empty handshape slot). This solution entails the assumption of two different signs in cases where an independent lexical base exists, though the lexical base looks the same as the bound root with the numeral handshape for “one.” Liddell (1996,

p. 208) explains these forms as having developed “through semantic extension,” e.g., the combined sign meaning “one week” extended its meaning to also mean “week.”

Out of these analyses, our own understanding of numeral incorporation most closely resembles Mathur and Rathmann (2011), as their analysis features a fully specified sign as a lexical base that is combined with a numeral handshake. Our reasoning for this will be discussed in Section 3.3.

2.3. Grammaticalization

Grammaticalization can be roughly described as a phenomenon of language change through which a linguistic item becomes more grammatical, e.g., a lexical content word like “go” becomes used in another function, e.g., as a future tense marker, and changes its behavior subsequently. However, when is a language change process also grammaticalization? Lehmann (1985) and Heine and Kuteva (2007) have devised some useful criteria, which we will refer to in detail in Section 5.2.

When it comes to theories of grammaticalization, the most prototypical example is the grammaticalization of a single element, as e.g., the emergence of a new morpheme that arises from the reanalysis of a linguistic construction. Reanalysis is a parameter of grammaticalization that is also emphasized by Hopper and Traugott (2003, pp. 50–63). Reanalysis describes a process of structural reinterpretation of an item in a given context. For example, the original structure of the “going-to” construction was a lexical verb (to go) with a to-infinitive sub-clause expressing an intention. The sentence “I am going to get groceries” in this original interpretation could likewise be expressed as “I am walking in order to get groceries.” The structure of the construction was then reanalyzed to instead consist of a temporal marker for future (be going to), followed by a lexical verb (“get,” in the above example). The reanalysis culminates in the usage in new contexts that do not allow for an interpretation in the original structure: “I am going to think about it” cannot be phrased as “I am walking in order to think about it,” as the latter sentence expresses a different meaning. As is apparent from this example, it is not only the individual item that undergoes a grammaticalization process, but rather an entire construction. Himmelmann (2004, p. 31) emphasizes this aspect of grammaticalization: “The unit to which grammaticization properly applies are constructions, not isolated lexical items.” The use of the grammaticalized item in new contexts is described as a “host-class expansion” by Himmelmann (2004, p. 32). That is, the grammatical item may combine with more and broader classes than in the source construction. Heine and Kuteva (2007, pp. 35–36) call this effect of a grammaticalization process the parameter of extension.

For languages with a long, written tradition, there is ample material to document and research instances of language change and grammaticalization. As a result, cross-linguistically similar processes have been attested that led to the identification of “clines” (Hopper and Traugott 2003, pp. 6–9), suggesting historical processes moving in a certain direction. The knowledge of these helps identify instances of grammaticalization where only synchronic data is available and internal reconstruction is the method of choice when grouping and classifying senses and functions of a linguistic item in question.

While phonological changes have been attested in sign languages, e.g., ASL (Frishberg 1975), and different historical stages of language use have been documented in historical dictionaries or historical descriptions of sign languages, these typically encompass single lexical signs, but not grammar or language in use. Even though historic records of sign language use are scarce, there are some findings of grammaticalization in sign languages (Pfau and Steinbach 2011; Janzen 2012). Pfau and Steinbach (2011, p. 684) note that the method of internal reconstruction is useful in sign language research, though less reliable than a diachronic comparative approach. Hypotheses can be built based on inferences from patterns of the synchronic grammar, assuming grammaticalization processes to be usually unidirectional and taking into account what is known about common grammaticalization paths (Pfau and Steinbach 2011, p. 684). Typically, grammaticalization in signed languages proceeds “from lexical element to free grammatical marker,” e.g., development of aspectual

and tense markers from verbs and adverbs, and less often or scarcely “from free grammatical marker to grammatical affix” (Pfau and Steinbach 2011, p. 689), which may be due to characteristics of signed language morphology, or the fact that sign languages tend to be relatively young languages (Janzen 2012, p. 836). In contrast to spoken languages, manual or facial gestures are a well-attested source of grammaticalization, even at times directly from gestures to functional elements, thereby skipping a step of lexicalization (Pfau and Steinbach 2011, pp. 689–93; Janzen 2012, pp. 829–34).

In the absence of historical data, synchronic language corpora become the database of choice. Representative, well-curated corpora allow for a thorough comparison of variant constructions. Following Bybee (2006, 2011), we regard the manifold manifestations of language use as a base for a reconstruction of cognitively organized rules that are informed by language use. Frequency of use is a relevant aspect for causing changes, e.g., frequency of co-occurrence of linguistic items as a prerequisite for chunking (Bybee 2011, pp. 70–71) most often leads at least to phonetic reduction. Phonetic reduction is a case of language change but not necessarily a grammaticalization process. The more frequently a construction is used, the more likely it is to become part of the cognitive linguistic rules; that is, the more likely it is to become entrenched (Bybee 2011, pp. 77–78). In the case of competing constructions, it is thus the one that is most frequent that is likely to become entrenched and, in consequence, lead to language change by abandonment of the other constructions, and “as long as frequency is on the rise, changes will move in a consistent direction” (Bybee 2011, p. 77). DGS lacks corpora with historic data of language use. However, a synchronic corpus has recently become available. Therefore, we can apply the apparent time hypothesis in our investigation of this potential grammaticalization process. According to Bailey et al. (1991), people within a generation tend to preserve the linguistic stage of their youth. Thus, a gradient age distribution in a language community can be taken to indicate language change. The oldest people in this approach represent the earliest language stage in the dataset, while the youngest people represent the most recent time period. This theoretical concept, used primarily in sociolinguistics in the wake of Labov (1963), attested its usefulness as an analytical tool through an evaluative test with real-time data (Bailey et al. 1991). The apparent time approach is the method of choice where historical linguistic data are lacking and has been used in sign language research e.g., by Hanke et al. (2017) and Dachkovsky (2022). Where possible, it should be complemented by diachronic, real-time approaches.

3. Materials and Methods

3.1. The DGS Corpus

The DGS corpus⁶ is compiled and curated by the DGS-Korpus project and serves as a reference corpus for a corpus-based dictionary of DGS (the DW-DGS). A total of 50 h of annotated and translated data are available to the public via three portals: one addressing the language community (*MY DGS*), and two addressing linguistic researchers (*MY DGS—annotated* and *MY DGS—ANNIS*) (see Hanke et al. 2020; Isard and Konrad 2022). For our study, we used the annotated parts of the reference corpus, which includes more data and more detailed annotations than the Public DGS Corpus.⁷

The DGS corpus is based around 560 h of natural and near-natural signing from 330 participants rooted in the deaf community. The participants use DGS fluently and on a daily basis. They were recorded in pairs throughout Germany from 2010 to 2012. Participants were asked to complete different tasks, such as conversations about given topics and retellings of stories. The corpus is balanced with regard to region, gender, and age of the participants (see Schulder et al. 2021). Four age groups were sampled based on the age of the participants at the time of the recording, ranging from the youngest group born between 1994 and 1981 (age group 18–30) to the oldest group born in 1950 and before (age group 61+). In between, there are age groups 31–45 (of people born between 1980 and 1966) and 46–60 (of people born between 1965 and 1951). At the time of conducting

our study in August 2022, the corpus data consisted of 92 h of lemmatized material with 671,851 tokens.

3.2. Corpus Annotation

The DGS data are annotated using the iLex database and annotation tool (Hanke and Storz 2008), which facilitates type-token matching and is searchable via SQL. Each type has a unique type id, but also an (id-)gloss name roughly hinting at the meaning of a sign.⁸ The annotation conventions (Konrad et al. 2022) specify all aspects of the annotation and provide an overview of different sign categories. We will briefly describe parts of the annotation process and the glossing conventions that are relevant to our study. The DGS-Korpus project is still ongoing, and the descriptions below match the conventions from the third release of the Public DGS Corpus.

In a first pass of annotation, translations are added on the sentence level; the signs are segmented and lemmatized (type-token matching); and the mouthings and mouth gestures⁹ are added (Hanke et al. 2023, p. 200). Any modifications and forms not exactly matching the citation forms are marked with the letter “a” (for “Abweichung”—deviation¹⁰). In the second step of annotation, these items are reviewed and categorized, possibly following a group discussion if the classification of the form is uncertain. As a result, information on the specific form of a given token is added, e.g., on a sign being one- or two-handed or a sign showing repeated movement (Hanke et al. 2023, p. 204). In preparation specifically for the study at hand, some of the annotations of numeral signs and numeral incorporating signs were revised by the annotation team.

During the lemmatization process, each sign token is matched with a type that has a unique, persistent ID, a gloss name, and a HamNoSys notation¹¹. The gloss name not only hints at the meaning of the sign but also includes a number to distinguish different signs with the same gloss name (e.g., lexical variants) and may include a letter to indicate a phonological variant (e.g., YEAR1A and YEAR1B). Numeral signs constitute their own category of signs and are marked by the prefix “\$NUM.” This prefix is attached to numerals of different kinds, such as ordinals, cardinals, divisions, etc., but also to some numeral incorporating signs, which do not have an independent base form (such as \$NUM-CLOCK1A)¹². In the DGS corpus, the numerals 1 to 10 are annotated using four different glosses: \$NUM-ONE-TO-TEN1A, \$NUM-ONE-TO-TEN1B, \$NUM-ONE-TO-TEN1C, and \$NUM-ONE-TO-TEN1D. These phonological variants differ with respect to hand orientation (away from the body/towards the body/one hand oriented in either direction). For our analysis and for readability in this paper, we subsumed and coded these types as \$NUM-ONE-TO-TEN.

Some of the forms deviating from the citation form of a sign type are annotated by adding qualifiers to the respective sign type gloss¹³. These qualifiers categorize and label recurring formational patterns of deviations from the assumed citation form across different signs that may be of phonetic, phonological, or morphological nature. Most relevant to this study is the qualifier “q” (quantification),¹⁴ which is added to either a numeral to specify the value of the numeral (possible values range from 0 to 10) or to a numeral incorporating sign to likewise specify the value of the incorporated numeral (possible values range from 1 to 10).¹⁵ The qualifier values may simply consist of a number but may also contain additional letters. The letters identify variants of the numeral handshape, e.g., the letter “d” in “1d” means a handshape with extended thumb, whereas “1” refers to the extended index finger.

For example, a non-incorporated form meaning “three weeks” looks like this in glosses: \$NUM-ONE-TO-TEN1A'q:3d WEEK1A. These refer to two distinct signs, as indicated by the blank space between the glosses (see also Figure 2a,b for the signs’ form). The second sign is WEEK1A, which matches its citation form and is, thus, not marked in any other way. The first sign is one of the phonological variants of the 1–10 numerals and is specified for the value “3d,” which means that the produced handshape is the 3-handshape involving the thumb. For comparison, the glossing of an incorporated form looks as follows:

WEEK1A'q:3d. Only the sign WEEK1A is articulated, and it is in this case the one specified by the value "3d," indicating that the sign is produced with the very same handshape as the numeral sign in the non-incorporated construction (see also Figure 3 for the sign's form).

All sign types (see Appendix D) and all examples (see Appendix C) can be found in the Public DGS Corpus. Note that the annotations in the Public DGS Corpus differ in some respects from the annotations of the reference corpus: Qualified forms are only indicated with an asterisk instead of a full specification of the modification, thus subsuming all qualified and deviating forms into one category. Signs of the category \$NUM, on the other hand, end by specifying the handshape according to the numbers 1–10.

3.3. Concept of Study

From the morpho-phonological complexity of the incorporating construction arises the question of how this structure may have developed and whether grammaticalization is involved. A balanced, annotated corpus with access to metadata information gives us the opportunity to take the synchronic variation, as well as sociolinguistic properties of participants such as region, gender, and age, into account. With a comparably large corpus, we have considerably more natural language data than was the case for any of the previous studies on numeral incorporation, some of which are based on elicited material, some on interviews, and some on both methods. This gives us the opportunity not only to attest the synchronic variation, but also to analyze it with respect to age groups, thus estimating trends of diachronic change according to the apparent time approach.

The annotated DGS corpus data suggest a three-partite variance of constructions when it comes to the combination of a numeral from 1 to 10 and a sign serving as a base sign. Additionally to constructions without incorporation, as in Figure 2a,b, and incorporated constructions, as in Figure 3, there is also an in-between structure that consists of an individually signed numeral and a base sign with numeral handshape incorporation (Figure 4).

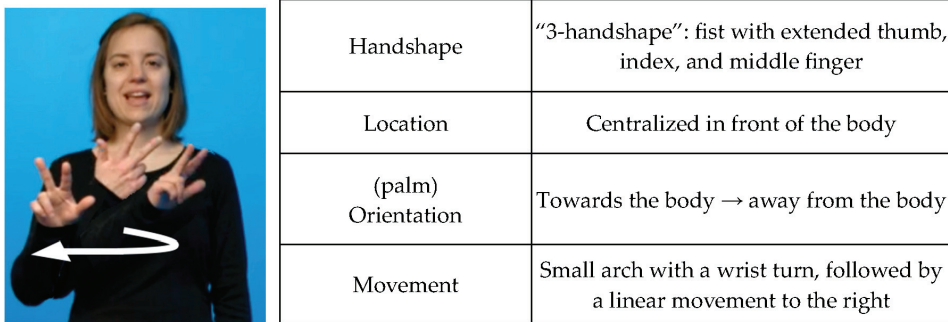


Figure 4. Phonological parameters and illustration of the cliticized form 'three weeks'.

In this construction, the handshape parameter of the lexical base sign is modified in a process of progressive handshape assimilation: The numeral's handshape perseveres throughout the subsequent lexical sign. Handshape assimilation is a regular and frequently occurring phonological process in sign languages (see Quer et al. 2017, pp. 53–54) and is often observed in quick or casual signing. Interestingly, an brief examination of the corpus data hints at a tendency towards regressive assimilation (handshape anticipation) in DGS rather than progressive assimilation. The assimilation to numeral handshapes thus stands out and indicates a morphosyntactic environment prone to cliticization. Additionally, there is a reduction of phonological material on part of the numeral sign. Though there are still two signs, they are less distinguishable as the numeral has lost its own movement value (slight forward movement), with the result of a smooth movement along both signs, resulting in a twist of the hand that comes from the change of hand orientation.

We will call the sequential production of two signs as shown in Figure 2a,b a **phrasal construction**, the incorporated form as shown in Figure 3 an affixation or **affixed form**, and the in-between construction we will call the **cliticized form** (Figure 4). It can be regarded as an incorporated form that is immediately preceded by the numeral whose value is identical to the incorporated one. Whenever we mention incorporating forms, we are referring to both affixed and cliticized forms. The existence of three variant constructions can be regarded as an indicator of ongoing language change, assuming that, synchronically, competing constructions show up during an ongoing language change process; for example a grammaticalization process.

Based on these observations, we propose a cline of three different constructions that show the progression towards grammaticalized numeral incorporation:

- **Phrasal construction:** In this construction the two elements are articulated independently from, but adjacent to, one another. The numeral keeps the location and orientation of its unmodified form, and the base is articulated with the handshape of its unmodified form. This is reflected in the corpus through the annotation of a numeral and a lexical base sign (see Appendix C Example (3) for an example featuring the phrasal construction).
- **Cliticization:** The numeral is still articulated as its own distinguishable sign with its own location and orientation. However, the noun loses its handshape and takes on the numeral handshape instead. The movement of the numeral sign is lost; instead, the movement of the lexical base is extended to accommodate the twist that is needed to match the hand orientations of both signs. This is reflected in the corpus through the annotation of a numeral and a modified lexical base sign (see Appendix C Example (4) for an example featuring the cliticized construction).
- **Affixation:** Here, the numeral only exists as the numeral handshape of the base sign. The two elements have become one sign. This is reflected in the corpus through the annotation of a single modified lexical base sign (see Appendix C Example (5) for an example featuring the affixed construction).

The existence of a cliticized form in the data speaks towards a process that starts with two independent signs, as proposed by Mathur and Rathmann (2011), rather than a bound morpheme and a bound root, as Liddell (1996) argues. For our study, we look at those instances of numeral incorporation that are based on independent base signs because we see the necessity of considering the relation between the independent sign and the incorporated form in order to explore the diachronic development of the incorporated form. This leaves open the issue of how to account for those incorporated forms that have no independent base sign, like the ASL sign O'CLOCK ("o'clock"). In DGS, there are also several signs without fully specified independent forms, including \$NUM-CLOCK1A "o'clock" (see Appendix C Example (6)). Synchronically, there is a syntactic paradigm with a suppletive form consisting of \$NUM-CLOCK1A, plus numeral incorporation for the numerals 1 to 10. From 11 onwards, signs such as CLOCK1 are used, which form phrasal constructions with any numeral. We will come back to the topic of numeral incorporations without a base sign, and their implications for the process as a whole in the conclusion (Section 6).

On a synchronic level, we agree with Mathur and Rathmann (2011) that the incorporated form is created through the combination of a base sign with all four phonological parameters and a fully specified numeral, and that, in this morphological process, the handshape parameter of the base sign gets deleted and is replaced by the handshape of the numeral. We call the form with complete incorporation "affixation" in the sense of "adding a morpheme to a structure" that is nevertheless being realized simultaneously with the base sign. The numeral morpheme within such an incorporated construction may furthermore be reanalyzed as consisting of a handshape only (as in Liddell 1996), and we hypothesize that, in fact, a process of reanalysis occurred diachronically. If we look at the numeral system of DGS, we see a model for such a reanalysis in the complex numerals. We suggest that these were reanalyzed as a meaningful handshape morpheme affixed to a meaningful

base morpheme consisting only of a centrally located, oriented movement, thus yielding numerals as 20, 30 to 90, or 100, 200, and up to 900, and so on, applying the strategy of combining a numeral handshape with a specific movement. Indeed, Sagara and Zeshan (2016, p. 29) call this strategy “numeral incorporation” because they put this simultaneous morphological process and numeral incorporation in lexical signs as described here in the same category (see also Zeshan et al. 2013, p. 363). We suggest that this view reflects cognitive processes (of reanalysis) of the language community at large. As a result of the reanalysis we propose, the numeral handshape morphemes “one of a X,” “two of a X,” and so on become available to be attached not only to elements in the numeral system to form complex numerals, but also to other lexical signs that denote typical quantifiable units. In other words, the numeral sign system becomes a model for a new set of constructions in which the numeral morpheme is affixed to other signs, such as temporal expressions.

3.4. Conducting the Study

At the preparative stage of our investigation, we deliberately excluded cases of attested numeral incorporation where no independent base sign can be found, as we were looking for signs that could occur in both incorporating and non-incorporating constructions. We also excluded all types that had fewer than five tokens with numeral incorporation in the corpus in order to avoid idiosyncrasies and other exceptional uses. A first query for signs meeting these two conditions gave us the following list of signs:

- WEEK1A, WEEK1B and WEEK1C (in the following subsumed and coded as WEEK1) “week.”
- MONTH1 “month.”
- YEAR1A and YEAR1B (subsumed and coded as YEAR1) “year.”
- DAY2 “day.”
- HOUR2A, HOUR2B, and HOUR2C (subsumed and coded as HOUR2) “hour.”
- EURO1 “euro” (currency).¹⁶
- OLD8B “age.”

Following our idea that numeral morphemes might have been abstracted from complex numerals, we included two incorporating signs from the numeral system, namely the hundreds and the thousands, as an assumed model for reanalysis and to check if they are used in the corpus as expected:

- \$NUM-HUNDREDS1 “hundreds.”
- \$NUM-THOUSANDS1 “thousands.”

These last two form a category we call “complex numerals,” while the longer list above is split into two semantic categories: the “temporals” (the first five items) and the remaining two signs that we group under “miscellaneous.”

The signs listed so far are the lexical hosts of our three constructions. In order to assess the frequency of complete numeral incorporation, we also need to look at the competing constructions: the phrasal construction and cliticization. As numeral incorporation occurs only with numerals ranging from 1 to 10 in DGS, we are only looking for occurrences with quantifying values in that range, be they quantified by a numeral sign or a numeral handshape. Our categorization for the annotation values “phrasal construction,” “cliticization,” and “affixation” is thus as follows:

1. Phrasal construction: a numeral from 1 to 10 followed by a lexical sign of the above list without the qualifier q.
2. Cliticization: a numeral from 1 to 10 followed by a lexical sign of the above list with qualifier q (specified for the same number value).
3. Affixation: a lexical sign of the above list with qualifier q and without a preceding numeral from 1 to 10.

You can see the corresponding constructions in glosses in Table 1.

Table 1. Annotation of numerals plus lexical sign for three construction types.

| Category | | Annotation Scheme and Example ‘Three Weeks’ |
|----------------------|---------|---|
| Phrasal Construction | Scheme | \$NUM-ONE-TO-TEN’q:n BASE-SIGN |
| | Example | \$NUM-ONE-TO-TEN’q:3d WEEK1 |
| Cliticization | Scheme | \$NUM-ONE-TO-TEN’q:n BASE-SIGN’q:n |
| | Example | \$NUM-ONE-TO-TEN’q:3d WEEK1’q:3d |
| Affixation | Scheme | BASE-SIGN’q:n |
| | Example | WEEK1’q:3d |

To find the sign tokens relevant for our study, we ran a query that selected all observations fitting the criteria described above and automatically annotated it according to our categorization. Furthermore, the query matched each observation with metadata on the participant producing the construction, e.g., the age group of the participant and their region. The result was 2992 observations of the signs listed above occurring in one of the three relevant constructions.

3.5. Hypotheses

The main prediction for our study is that numeral incorporation is the result of a language change process, specifically a grammaticalization process. The property of language change that we will be testing on the basis of corpus data is the progressive change in language usage throughout time. Following the apparent time hypothesis (see also Section 2.3; Bailey et al. 1991), we will test this by looking at the different age groups represented in the corpus and comparing their language use. Furthermore, we use the cline of phrasal construction, cliticization, and affixation as a basis of looking at the progressive change in form and will be comparing the usage of each construction by age group.

The idea of this perspective on language change is that the newly developing form, in this case, the incorporated form, is used more by younger people than by older people (see Hypothesis 1).

Hypothesis 1. *The older participants are, the less likely they are to use incorporated constructions (clitics and affixes).*

We hypothesize that complex numerals might be the original source of the numeral handshape as a unit of meaning. While we cannot test this directly without diachronic data, this presumption implies that numerals exclusively or almost exclusively occur in incorporated constructions (see Hypothesis 2).

Hypothesis 2. *Complex numerals have a (very) strong tendency to be used in incorporated constructions.*

Temporals are the signs that are the most well-known for their tendency to incorporate numeral handshapes. We hypothesize them to also show a tendency to occur with incorporations, but we expect this tendency to be weaker than in complex numerals (see Hypothesis 3).

Hypothesis 3. *Temporals have a fairly strong tendency to be used in incorporated constructions.*

The signs that we summarize in the “miscellaneous” group are signs that we think are just starting to show incorporations. Thus, we expect them to still prefer use in phrasal constructions (perhaps even strongly prefer) (see Hypothesis 4).

Hypothesis 4. *Miscellaneous signs show some incorporations but overall tend to be used in phrasal constructions.*

In summary, according to sign group, we expect a decreasing amount of incorporation in this order: complex numerals; temporals; and miscellaneous. With regard to the age group, we expect more incorporations the younger people are.

3.6. Analysis

The analysis of the data was completed in R Studio using R version 4.2.1, the tidyverse package (Wickham et al. 2019), the lme4 package (Bates et al. 2015), and the car package (Fox and Weisberg 2019). A generalized linear mixed model was fitted to the data using the glmer function. The outcome variable was the categorization of the construction, a categorical variable with three levels: phrase, clitic, and affix. In order to fit the model, the variable was dummy coded using the levels 0, 1, and 2 (respectively corresponding to phrase, clitic, and affix). The outcome is discrete count data with three different levels, calling for a binomial model.

The model included two predictors: sign group and age group. As described above, the signs in question were grouped into three categories based on their semantics: complex numerals, temporals, and miscellaneous. This grouping is a categorical variable with three different levels. The second predictor is the age of the participant. As age is represented in the corpus in terms of age group, this is likewise a categorical variable with four levels: 18–30, 31–45, 46–60, and 61+.

Two random effects were added to the model: sign gloss and participant. Sign gloss was added in case some signs show particularly strong tendencies towards (non-)incorporation and participants were added in case some people have particularly strong preferences for (non-)incorporation. Sign gloss is a categorical variable with nine different levels, consisting of the nine incorporating signs we selected for this study. Participant is a categorical variable with 305 levels, consisting of the participant IDs of all participants in the dataset.

The resulting model formula is **Categorization ~ 1 + Sign Group + Age Group + (1 | Sign Gloss) + (1 | Participant)**.

Originally, region was also supposed to be included in the model as a random effect. Unfortunately, there were convergence issues when more than two random effects were included. Instead, we ran a model with region as a fixed effect to test whether regionality has an effect on use of incorporation (Categorization ~ 1 + Region + (1 | Sign Gloss) + (1 | Sign Group)). The results show non-significant slope estimates for all regions except for one: Leipzig has a little more incorporation than the other regions with an estimate of 1.1862 and a *p*-value of 0.0488. A post-hoc ANOVA on the model shows that region is not a significant factor overall (*p* = 0.1451). Thus, region does not seem to have a systematic influence on the use of incorporated forms and may be excluded.

4. Results

The data collection of occurrences of the nine selected signs resulted in a total of 2992 observations. For an overview of the absolute sign frequencies per age group, see Table A2 in Appendix E. In this section, the results of the fitted generalized linear mixed model will be shown.

With regard to sign group, the model came up with the estimates shown in Table 2.

Table 2. Results showing the influence of the sign’s group.

| | β | Standard Deviation | z-Value | p-Value |
|----------------------|---------|--------------------|---------|-------------------------------|
| Intercept | 7.8104 | 1.2794 | 6.105 | $p < 0.001$ *** ¹⁷ |
| Temporals | −1.8503 | 1.2491 | −1.481 | 0.13853 |
| Miscellaneous | −8.4218 | 1.5581 | −5.405 | $p < 0.001$ *** |

The intercept shows the level of incorporation for signs in the complex numerals group used by 18–30 year olds. As expected, this is a rather high value, and the slope estimates for temporals and signs in the miscellaneous group show a decreasing tendency regarding incorporations. For the latter group in particular, this tendency is very strong. This indicates that the most incorporations are produced by 18–30 year olds in complex numerals. Temporals show slightly less incorporation than complex numerals as shown by the negative estimate, but the difference is not strong enough to be significant. The two signs in the miscellaneous groups show a lot less incorporation than complex numerals, and this difference is very highly significant.

The estimates for the predictor age group are shown in Table 3.

Table 3. Results showing the influence of the participants’ age group.

| | β | Standard Deviation | z-Value | p-Value |
|------------------|---------|--------------------|---------|-----------------|
| Intercept | 7.8104 | 1.2794 | 6.105 | $p < 0.001$ *** |
| 31–45 | −1.4134 | 0.6745 | −2.095 | 0.03613 * |
| 46–60 | −1.8799 | 0.6819 | −2.757 | 0.00584 ** |
| 61+ | −2.9070 | 0.6982 | −4.163 | $p < 0.001$ *** |

As expected, all slope estimates are negative, meaning that the reference group of 18–30 year olds is the most likely to use incorporated forms. Those who are 31–45 and 46–60 years old are significantly less likely to use incorporated forms than 18–30 year olds. Participants in the 61+ age group are significantly less likely to use incorporated forms than 18–30 year olds.

A post-hoc ANOVA confirms that both factors have significant effects on the use of numeral incorporation overall (see Table 4).

Table 4. Results of a post-hoc ANOVA confirming the significance of sign group and age group.

| | χ^2 | Degrees of Freedom | Pr (>Chisq) |
|-------------------|----------|--------------------|-----------------|
| Sign Group | 33.845 | 2 | $p < 0.001$ *** |
| Age Group | 19.135 | 3 | $p < 0.001$ *** |

The estimated marginal means for incorporation according to the model can be seen in Figure 5. These allow for a clear comparison between the different categories.

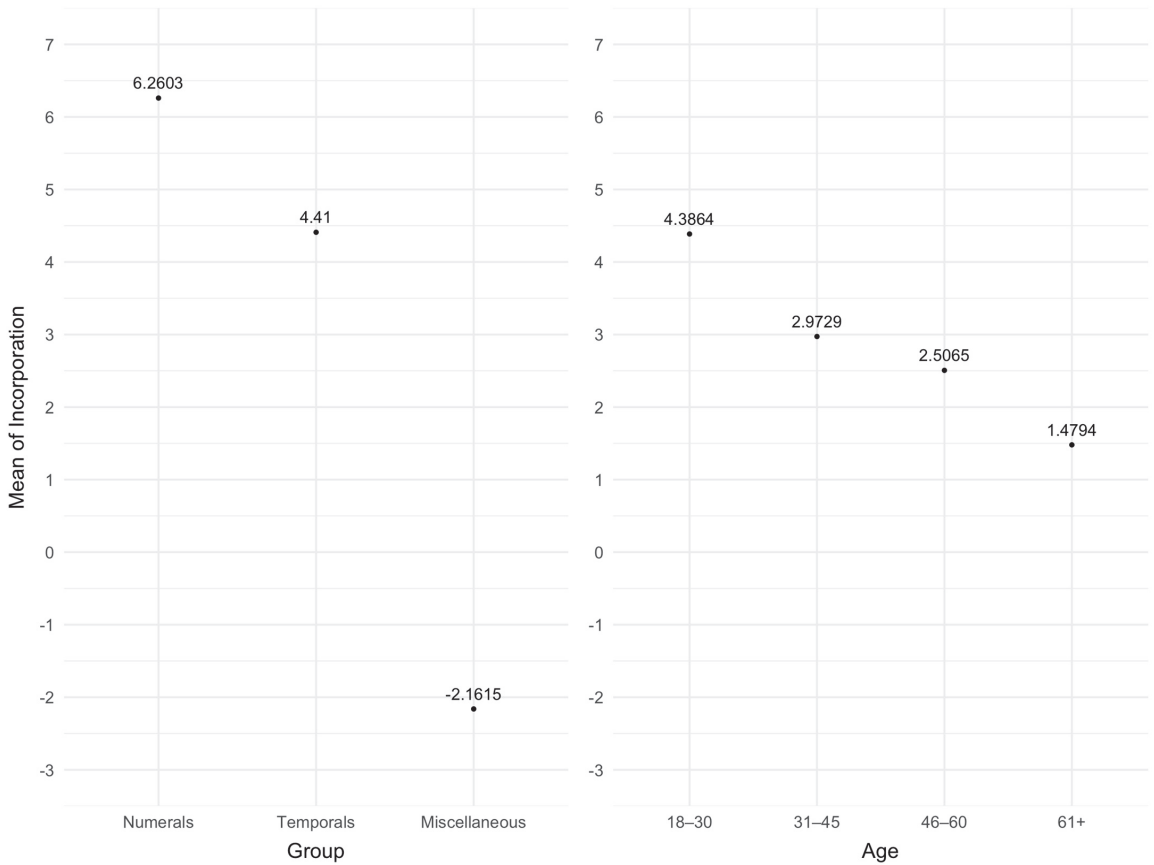


Figure 5. Means of Incorporation as calculated by the linear regression model.

The means show that the sign group has a strong effect on the use of incorporated forms with complex numerals having the highest mean (6.2603), and miscellaneous signs having by far the lowest mean (−2.1615). This overall pattern is expected, though we anticipated the difference between complex numerals and temporals to be bigger, as we suggested complex numerals as a potential origin of numeral incorporation. The miscellaneous signs match our expectations as possible future host class expansions that are just starting to be used. Concerning the age of participants, we see the expected pattern of younger people using more numeral incorporations (4.3864) than older people (1.4794). The two middle groups fit the overall pattern, with the mean of the 46–60 group being slightly lower (2.5065) than that of the 31–45 group (2.9729). However, the means of these two groups are very close to each other, so it seems that they use numeral incorporation in similar amounts.

The relative frequencies of the different constructions per age group and per sign group can be seen in Figure 6.

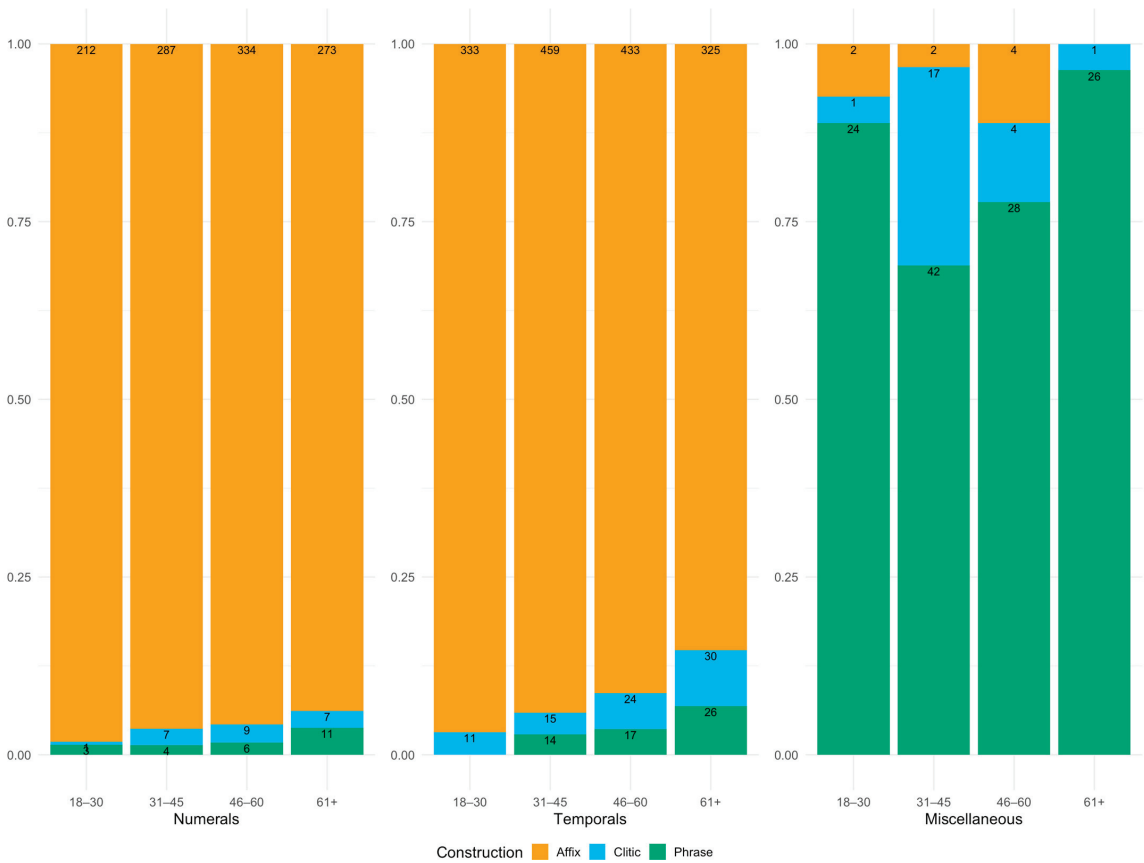


Figure 6. Construction frequencies per age group and sign group.

The visual difference between the first two groups and the third group is striking: While complex numerals and temporals are largely incorporated and only contain some phrasal constructions, the two signs in the miscellaneous group are used in phrasal constructions much more than in incorporated forms, though uses of affixes and clitics still occur.

Regarding the age groups, complex numerals and temporals again show very similar patterns: the youngest age group shows the strongest preference for incorporated forms with barely any phrasal constructions. As the participants get older, the preference for affixed constructions grows weaker (though all groups show that same preference for affixes), and the use of both clitic and phrasal constructions gradually increases. In the miscellaneous group, all age groups show a preference for phrasal constructions. This preference is the weakest in the age group including 31–45 year olds at just under 70% phrasal constructions. The oldest age group shows the strongest preference for phrases with only a single occurrence of an incorporated form, which is a cliticization. As the youngest age group shows the second-strongest preference for phrasal constructions, there is no consistent pattern regarding the usage by different age groups in the miscellaneous group.

It is possible that individual signs show singular usage patterns. The usage of each sign per age group is visualized in Figure 7.

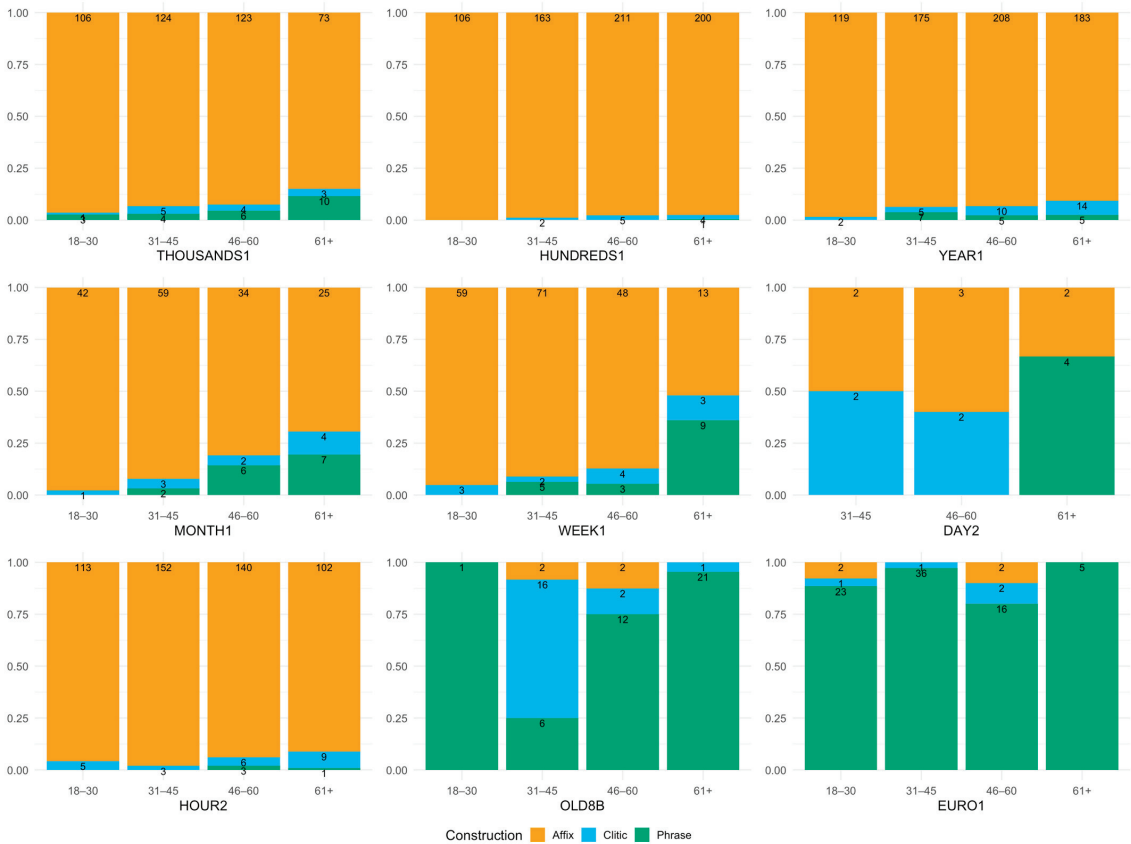


Figure 7. Construction frequencies usage per sign.

The two complex numerals, \$NUM-THOUSANDS1 and \$NUM-HUNDREDS1, behave rather differently from each other, with THOUSANDS containing quite a few phrasal constructions and HUNDREDS containing only one in the oldest age group. All temporals, except DAY2 (that is YEAR1, MONTH1, WEEK1, and HOUR2), show the expected general pattern of decreasing use of incorporations with increasing age. DAY2 does not follow this pattern beyond only having phrasal constructions in the oldest age group. It should be noted that this sign has by far the fewest occurrences overall at only 15 total tokens (for comparison: the sign with the second-fewest tokens is OLD8B at 63 tokens). The signs grouped as miscellaneous, OLD8B and EURO1, both show strong preferences for phrasal constructions in the oldest age group. Additionally, OLD8B contains exclusively phrasal constructions in the youngest age group, though this intersection consists of only one token. For EURO1, the age group of 31–45 year olds also shows a strong preference for phrasal constructions, while the youngest age group and the 46–60 age group both contain some, though few, incorporated tokens. OLD8B, when considered without the one token of the youngest age group, actually shows the same pattern as the temporals: a decreasing use of incorporations with increasing age though the temporals have a general preference towards affixes, while OLD8B has a general preference towards phrases. Nonetheless, the overall pattern and the high number of clitics in the 31–45 age group, in particular, is consistent with the expected usage pattern.

5. Discussion

5.1. Hypotheses Evaluation

The results showed how participants of different ages use three categories of signs with regards to numeral incorporation, including three levels of incorporation: phrasal (not incorporated), clitic (partially incorporated), and affix (completely incorporated). Keep in mind that our dataset only contains quantified uses from 1–10 of the signs in question, so all of our frequency observations are made upon this basis. The results of our study will now be evaluated with regards to the hypotheses, repeated below for convenience.

Hypothesis 1. *The older participants are, the less likely they are to use incorporated constructions (clitics and affixes).*

Hypothesis 2. *Complex numerals have a (very) strong tendency to be used in incorporated constructions.*

Hypothesis 3. *Temporals have a fairly strong tendency to be used in incorporated constructions.*

Hypothesis 4. *Miscellaneous signs show some incorporations but overall tend to be used in phrasal constructions.*

The correlation between participants’ age and use of incorporations, which is described in the first hypothesis, was confirmed by the data. The oldest and the youngest age groups do indeed bookend the progression with the fewest and the most uses of incorporations, respectively. The two middle groups, likewise, fit the expected pattern with the group of 31–45 year olds using more incorporations than the group of 46–60 year olds. The graduality of this change is underlined by the fact that none of the differences between the adjacent age groups are significant (see Table 5), though the overall effect of age is highly significant (see Table 4 in Section 4). According to the apparent time approach, this correlation between age and usage frequency indicates language change. In our case, the change is from the usage of only phrasal constructions over cliticizations towards the use of affixes. Based on the use of phrasal constructions by older people, it seems that this construction was more common in an early stage of DGS, or it may have been the only possible construction to express this meaning. The gradual change throughout the middle age groups indicates that DGS progressed to include increasingly more incorporations over time. The youngest age group uses exclusively incorporated forms for temporals and very few phrasal constructions for complex numerals, indicating that in the most contemporary forms of DGS, the incorporations may be the only form remaining in use.

Table 5. Results of a paired post-hoc test on age group.

| | β | Standard Error | Degrees of Freedom | z.Ratio | p.Value |
|-----------------|---------|----------------|--------------------|---------|---------|
| (18–30)–(31–45) | 1.413 | 0.675 | Inf | 2.095 | 0.1545 |
| (31–45)–(46–60) | 0.466 | 0.500 | Inf | 0.933 | 0.7873 |
| (46–60)–61+ | 1.027 | 0.497 | Inf | 2.066 | 0.1644 |

The complex numerals show the same age progression, with older people using the most phrasal constructions and young people using the least. With a total of 24 phrasal constructions in the complex numerals group (0.0208%), we do indeed observe the tendency towards incorporated forms as predicted by the second hypothesis. However, the tendency is not as strong as expected. In Section 3.3, we suggested that the complex numerals may be the origin of numeral incorporation and the grounds of a reanalysis of the numeral handshake as an affix. If this were true, we should be observing almost exclusive usage of incorporated forms, or at least a significantly lower percentage of phrasal constructions

than in the temporals group. With 0.0208% phrasal constructions in the complex numerals and 0.0338% phrasal construction in the temporals, this is not the case. The model estimates presented in Table 2 in Section 4 confirm that the incorporation usage in complex numerals does not significantly differ from the usage in temporals. Thus, while we can technically confirm our hypothesis of strong tendencies towards incorporation in complex numerals, we do not come to the expected conclusion of complex numerals as the origin for numeral incorporation.

Another interesting observation when looking at the complex numerals is the difference between the individual signs. Almost all of the phrasal constructions in the complex numerals occur with \$NUM-THOUSANDS1, \$NUM-HUNDREDS1 is only once used in a phrasal construction. A closer look at this singular token reveals that this is actually not exactly the construction we were looking for. The intended meaning of the construction is the year 1940, which would usually be signed \$NUM-TEEN1'q:9 \$NUM-HUNDREDS1 \$NUM-TENS1'q:4 "nineteen-hundred-forty," which would not have appeared in our data due to the lack of a \$NUM-ONE-TO-TEN gloss. However, in the given observation, the signer signs \$NUM-ONE-TO-TEN1A'q:9 \$NUM-HUNDREDS1 \$NUM-TENS1'q:4 "nine-hundred-forty." It is questionable whether this observation should have been part of our data. However, since we did not manually check all of the 2992 observations compiled in the automatic annotation-based data selection process, such deviations in the data could not be detected and, therefore, not be excluded. Disregarding this singular phrasal usage with \$NUM-HUNDREDS1 makes the difference between the two complex numerals even more drastic, as it is then only \$NUM-THOUSANDS1 that is used in phrasal constructions at all. We cannot explain this difference on the basis of our data. On a phonological level, the two are so similar that this seems unlikely to be the cause. One possibility is that \$NUM-HUNDREDS1 is used more frequently overall (see Table A2 in Appendix E) with 692 total appearances in our dataset versus 462 for \$NUM-THOUSANDS1. If numeral incorporation is still emerging in complex numerals as well, it is plausible that it would first be established in the more frequent numerals and then spread to the less frequent ones. This, again, speaks against complex numerals as the origin of numeral incorporation.

The group of temporals, which is our main testing ground for incorporation usage, shows the expected pattern of younger people using more incorporated forms than older people. However, the overall tendency of temporals towards affixation in particular is stronger than we would have expected with 1550 affixations, 80 cliticizations, and 57 phrasal constructions in total. Even in the oldest age group, 93% of temporals have an incorporated numeral handshape (85% affixations). This indicates that the emergence process has likely started long before the oldest generation represented in our corpus and has progressed quite far. This is also supported by the fact that the younger generation not only does not prefer the phrasal construction, but in fact does not use it at all, as all 344 tokens show some kind of incorporation. The temporals' pattern also supports our claim of the cliticization construction as an in-between stage in the emergence of the numeral affix, as the use of the clitic, just like that of the phrasal construction, is the strongest in the oldest age group and the weakest in the youngest. Overall, the pattern shown here is in line with our hypothesis of numeral handshapes emerging through a language change process.

The last group of signs, termed the "miscellaneous" group, shows no clear overall pattern. However, with 21% total incorporating tokens, we can say that the signs certainly can be used in incorporating constructions. Yet the preference for phrasal constructions is very clear in all age groups. A look at the individual signs shows that EURO1 has only a few incorporations. None of the incorporations are in the oldest age group, but there is no overall pattern visible. OLD8B, in contrast, is used in a pattern more fitting to our expectations. Disregarding the youngest age group, for which there is only one token of this sign, the expected pattern of more incorporations in the second youngest age group and the lowest number of incorporations in the oldest age group can be found once more. However, a closer examination of these tokens reveals that 13 of the 23 incorporations were produced by the same participant, and that 18 of them are from the same region (Cologne).

The latter is less surprising as OLD8B is a regional sign used in the west of Germany, but the fact that over half of the tokens are from the same participant does skew our data. We will thus refrain from drawing conclusions beyond the possibility of incorporations for both signs. Hypothesis 4 is thus supported, but our claims regarding usage in different age groups cannot be verified in this group.

Besides the results regarding these hypotheses, we also observed that the phonological constraints regarding the base sign in DGS described by Mathur and Rathmann (2011) hold in our data for the most part but are not as strict as previously described. While all of the temporals that we considered (YEAR1, MONTH1, WEEK1, DAY2, and HOUR2) use the 1-handshape (extended index) in their unmodified form, we were also able to observe incorporations in OLD8B and EURO1, neither of which has a 1-handshape. Their handshapes are still rather unmarked, but nonetheless, these uses may indicate a shift in the phonological constraints. The 1-handshape constraint still holds for the temporals, however. For WEEK1 and DAY2, we also looked at lexical variants that do not have the 1-handshape: WEEK2 (46 tokens) and DAY1A/DAY1B (157 tokens). Neither of these show incorporating tendencies.

In conclusion, the data have shown that age groups do in fact differ in the way they use numeral incorporation, and that the resulting pattern supports a language change hypothesis. More research is needed both in regard to the usage of numeral incorporation in temporal signs to see whether it will go on to progress in a way consistent with language change, and also with regard to our miscellaneous group with signs that may incorporate more in the future. That being said, our main hypothesis was not only that numeral incorporation is the result of language change, but specifically the result of grammaticalization. In the following, we will discuss to what extent the properties of numeral incorporation are consistent with a grammaticalization analysis.

5.2. Numeral Incorporation as Grammaticalization

The phenomenon known as grammaticalization is associated with various different steps the construction in question typically goes through. Here, we will look at the parameters suggested by Lehmann (1985) and Heine and Kuteva (2007) and discuss whether they apply to numeral incorporation. We will first list the parameters and then describe the properties of numeral incorporation with respect to these parameters. Keep in mind that some of the categories that Lehmann, Heine, and Kuteva use to describe their parameters are either not transferrable to sign languages or may not have been researched in sign languages yet, so that their status is unclear.

Lehmann's (1985) grammaticalization parameters:

1. Attrition: loss of semantic, phonological, and morphological properties.
2. Paradigmaticization: emergence of morphological paradigms.
3. Obligatorification: the new construction becomes obligatory, alternatives disappear.
4. Condensation: decrease in the element's scope.
5. Coalescence: increase in bondedness.
6. Fixation: loss of variation in positioning.

Heine and Kuteva's (2007) grammaticalization parameters:

1. Reinterpretation: item is reinterpreted in a given context.
2. Extension: item can be used in new contexts.
3. Desemanticization/semantic bleaching: meaning components that are not compatible with the new usage are lost.
4. Decategorialization: loss of morphosyntactic properties.
5. Erosion: Loss of phonetic substance.

With regards to **reinterpretation**, we originally suggested complex numerals as the origin of numeral incorporation with their handshape being reinterpreted as an affix rather than just a phonological parameter. As described in Section 5.1 above, our data does not support this hypothesis and we, consequently, cannot discern a reinterpretation process

in the emergence of numeral incorporation. Nonetheless, we argue that an **extension** has taken place, or rather, is currently taking place as OLD8B and EURO1 are starting to be used in incorporated constructions. Furthermore, we suggested a cline for the emergence of numeral incorporation, starting with the phrasal construction featuring two fully specified signs: the cliticized construction in which the signs begin to merge, and, finally, the affixed construction in which the two signs have become one sign. Compared to the fully specified numerals in the phrasal constructions, the numeral handshapes in the incorporated constructions show **phonological attrition/erosion** (loss of three of their four phonological parameters) and **morphological attrition** (change from a fully specified independent sign and free morpheme to an affix and bound morpheme). The base sign likewise undergoes a phonological reduction as its handshape is deleted. Due to phonological constraints on incorporation, not all numeral signs can be reduced to numeral handshapes: morphologically complex numerals such as \$NUM-TENS2A'q:4 "forty" feature a movement that is integral to the meaning of the sign and can thus not be reduced to their handshape. As a consequence of these constraints, a **closed paradigm** of numeral handshapes, ranging from 1 to 10, emerges.

Due to the simultaneous nature of the incorporation, the bond (**coalescence**) between the numeral handshape and the base sign is very strong, as they become one sign. In Example (7) (see Appendix C), we observe a very interesting construction: HOUR2B'q:3d HALF6 "three-and-a-half hours." The numeral handshape for "three" is incorporated into the sign HOUR2B. The sign HALF6, which semantically belongs to "three," and would have immediately followed the numeral sign in a phrasal construction (as we can see in several other constructions), cannot come between the numeral handshape and the base sign here, confirming their bondedness at least on a phonological level. We have found at least four observations like this (all with a sign meaning "half"), produced by four different participants. Exploring the exact nature of this construction on a morphological level is beyond the scope of this paper, but the structural separation of "three" and "half" is quite interesting and would be a great topic of study (perhaps combined with an exploration of whether the incorporating base sign can still be modified, that is, whether constructions such as BEAUTIFUL1A WEEK1'q:3 "three beautiful weeks" are possible).

Parameters that do not seem to fit for numeral incorporation at all are the ones referring to a reduction in semantic content (**semantic attrition/desemanticization**). Numerals are fairly simple in their semantics to begin with, and there does not seem to be any meaningful difference between the phrasal \$NUM-ONE-TO-TEN1A'q:3 WEEK1 and the affixed WEEK1'q:3. As we did not include a semantic analysis in our study, we cannot rule out minor semantic differences, but as it stands now, we see no shift in meaning.

The parameter of **obligatorification** may not apply yet, as there is the alternative of using the phrasal construction. However, the fact that the participants in the youngest age group are using exclusively incorporated forms for temporals may be an indicator for an ongoing obligatorification process that will be completed in the future.

The parameters of **decategorialization** and **condensation** (reduced scope) are difficult to judge for numeral incorporation, as the morphosyntactic properties of signs have not been researched enough to describe them for the phrasal construction, nor for the affixed construction. Regarding the morphological properties of incorporated forms, we can see in our data that the incorporated constructions can be modified further to mark morphological categories such as regularity ("every three weeks"). However, as this category is specific to temporal expressions, the modification does not indicate a general consideration of the incorporated construction as a single noun with the same morphosyntactic properties as other nouns. A conclusion regarding a decategorialization of the construction cannot be made.

Lastly, the parameter of **fixation**, described by Lehmann (1985) as the loss of variation in positioning, cannot be applied to this construction due to the simultaneity of the affix and the root. The phrasal construction tends to occur in the order of the numeral, followed by the base sign. Since the affix is a piece of simultaneous morphology, its position to

the root cannot be described as preceding or following and thus cannot match that of the numeral in the phrasal construction.

In conclusion, we believe that too many essential components of a grammaticalization process are missing to claim that numeral incorporation is the result of a grammaticalization process. In particular, the lack of semantic bleaching/desemantization on the one hand, and the fact that the complex numerals considered here do not hold up as a possible base for reanalysis on the other hand, indicate that we are not dealing with a grammaticalization process.

5.3. Alternative Approaches

Having concluded that numeral incorporation is in an ongoing process of language change but does not seem to be an instance of grammaticalization, the question arises if there might be another known language change process that might be a better fit for the description of numeral incorporation. In this section, we want to discuss some alternatives in the light of our study's results. Approaches described in this section include constructionalization, lexicalization, chunking, and compounding. Some of these may be treated as parts of a larger grammaticalization process (e.g., chunking) or as a larger language change category, under which grammaticalization can be subsumed (e.g., constructionalization). However, as all of these processes can also occur without being classified as grammaticalization, we believe them to be worthy of discussion here as alternatives.

Traugott and Trousdale (2013) describe **constructionalization** as the creation of a new form–meaning pair with an arbitrary association between form and meaning (p. 1). In the case of two pre-existing elements forming a new construction, they clarify that the result exhibits changes in syntax, morphology, and meaning, e.g., the meaning of “cupboard” is non-compositional as it cannot fully be derived from its components “cup” and “board.” This is similar to the “semantic bleaching” parameter in grammaticalization processes and, as already described, this does not apply to numeral incorporation. Furthermore, Traugott and Trousdale (2013) discuss clippings and blends as lexical constructionalizations (p. 150). The forms resulting from clippings and blends do not look entirely dissimilar to numeral incorporations, as they likewise combine some of the phonological aspects of one element with some of the phonological aspects of another element, e.g., the word “brunch” combining the “br” from “breakfast” with the “unch” from “lunch.” However, as Traugott and Trousdale point out, clippings and blends do not arise gradually (p. 150). This stands in direct contradiction to the findings of our study, as numeral incorporation has developed gradually both in terms of the temporal development (as younger age groups use more incorporation than older age groups) and in terms of the form gradually emerging (shown by the phrase—clitic—affix cline). Thus, constructionalization cannot account for all of the aspects of (the emergence of) numeral incorporation.

Lexicalization, as it has been described for sign languages so far, likewise includes a shift in meaning (Frishberg 1975; Liddell and Johnson 1986). It furthermore assumes that each newly lexicalized form is stored as a single unit in the lexicon (Liddell and Johnson 1986) and, thus, excludes productive word formations. However, as we have shown, numeral incorporation is certainly productive.

Chunking describes the phenomenon of “units or word strings that are often produced together (. . .) becom[ing] units or chunks of their own right” (Bybee 2011, p. 2). The new chunks are then stored and processed together. The prerequisite of the units being often produced together certainly fits numeral incorporation: the signs that allow for incorporation are ones that are commonly quantified and are consequently often produced with a numeral. However, Bybee (2011, p. 3) also points out that the meaning of the chunks usually becomes non-compositional with time, and that the association with the original components fades until they are not activated alongside the chunk anymore. This is certainly not the case for numeral incorporation, as the meaning is very much compositional, and the association with the original number and base sign is not only given but also unlikely to fade (given the compositionality of the meaning).

Lastly, we want to consider the possibility of numeral incorporation being a case of **compounding**, as Meir (2012) and Jones (2013) suggest. In this scenario, our description of the numeral handshape as an affix would be incorrect and the numeral handshape would instead be considered a modifier with the base lexical sign acting as the head of the compound. Meir (2012, p. 101) describes two types of compounds in sign languages, the second of which “combines certain phonological parameters from two different sources to create a single sign,” and names numeral incorporation as one of the instances of this type of compounding, though she acknowledges that this is a rather unusual analysis (Meir 2012, p. 101). However, when describing affixes in sign languages, which are usually simultaneous, she adds that compounds and affixes are difficult to distinguish in sign languages (Meir 2012, p. 103). Her two criteria for differentiation are (i) the number of bases that a given element can combine with, as affixes should be more productive and combine with more bases than compound elements, and (ii) allomorphy, which, according to Meir (2012, p. 103), is more common in affixes. Regarding the productivity of numeral handshapes, we are uncertain of a reliable judgement. Our dataset included nine different signs, but including incorporations without an independent base sign would result in many more signs being added to the list of incorporating signs (e.g., \$NUM-CLOCK1A, \$NUM-GRADE1, and \$NUM-FROM-TO1, just to name a few). Furthermore, we have shown through the inclusion of OLD8B and EURO1 in our study that numeral incorporation has progressively been attached to more and more bases. Furthermore, the fact that we are not looking at a single element but a full paradigm of numeral handshapes from 1 to 10 does make the form seem more productive overall. Still, the numeral handshapes can certainly not compete with Meir’s (2012, p. 103) example of the affix “-ness” which can attach to about 3058 English words. Regarding the matter of allomorphs in numeral incorporation, we have shown handshape variations in multiple numeral signs, such as “one,” being articulated with either the thumb or the index finger (see also the overview in Appendix B Figure A2). As all of these handshapes can be incorporated, the case for multiple allomorphs for one number affix can be made. Thus, we stick to our analysis of the incorporated handshape as an affix.

In conclusion, none of the analyses described here can truly capture the numeral incorporation phenomenon in all of its aspects. This may be due to the lack of in-depth understanding of language change in sign languages overall, or it may be an indicator that this is a new category of language change.

6. Conclusions

Our corpus-based study reveals that there is good reason to assume that numeral incorporation is the result of a language change process, though likely not a grammaticalization process. The synchronic co-existence of three different constructions (phrasal, cliticized, and affixed) that are used in varying amounts by different age groups shows that change is still ongoing. This language change process lacks some of the essential properties of grammaticalization, though none of the other categorizations discussed here seem to be a good fit, either.

The observed tendencies towards incorporation in temporals suggest that incorporation in quantified temporal signs may become obligatory in the future, especially given the use of exclusively incorporated forms in the youngest age group. In the complex numeral \$NUM-HUNDREDS1, we already find exclusive incorporation indicating obligatoriness of the numeral handshape, confirming that an obligatorification is possible. We do not find an obligatory use of incorporated forms in \$NUM-THOUSANDS1, however. The reason for this difference could lie in the difference in overall frequency of the two signs, though this matter deserves further investigation.

With regards to the incorporating signs without a base sign such as \$NUM-CLOCK1A (described in Section 3.3), it is possible that these represent an ending point to the emergence of numeral affixes. Given the obligatorification tendencies we can see in our data, it is possible that \$NUM-CLOCK1A originally did have a base sign that disappeared after

the incorporated construction became obligatory. Presumably, this would only happen in signs that have lexical variants that are used in non-quantified contexts and in quantified contexts surpassing an amount of 10 (such as CLOCK1), as the remaining use in these contexts would otherwise prevent the disappearance of the base sign. Considering the base signs included in our study, these are thus unlikely to disappear, as they are also used in non-quantified contexts. Still, this interpretation is a possible explanation for incorporated forms without a base sign.

Numeral incorporation in temporals seems to depend on several factors, including phonological and semantic constraints. However, as is common with corpus data, we see a lot of variation in use. Strict phonological constraints do not predict individual signers building forms such as EURO1'q:3d, thus violating the principle of handshape markedness. Though these forms are unexpected, conversation partners understand them, and it is imaginable that, one day, given an increased frequency of these forms, a new member has entered the class of host signs for numeral incorporation, overriding any previous phonological restrictions.

In the future, tendencies observed here, such as the incorporations in non-temporal signs and the beginning obligatorification of the incorporating construction, deserve further observation. When more contemporary data become available in the future, it can be used for comparison to our dataset, thus allowing for a diachronic comparison of language use throughout time. Our corpus-based results could, furthermore, be supplemented by an experimental approach to the subject, e.g., through an acceptability judgement task asking participants of different age groups to judge incorporation and phrasal constructions.

In conclusion, we have found that numeral incorporation seems to be the result of a still ongoing process of language change, though this claim cannot be made with certainty without diachronic data. While we do not believe this process to be an instance of grammaticalization, operating within the framework of grammaticalization has allowed us to describe many aspects of the usage of numeral incorporation in interesting and useful ways. It is our hope that future research will be able to discover more about the exact nature of the emergence of numeral incorporation and diachronic language change in sign languages overall.

Author Contributions: Conceptualization, F.O., A.M., S.W., G.L.; methodology, F.O., A.M., S.W., G.L.; formal analysis, F.O.; resources, F.O., A.M.; SQL queries, G.L., A.M.; writing—original draft preparation, F.O., A.M.; writing—review and editing, F.O., A.M., S.W., G.L.; visualization, F.O., S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained by the DGS-Korpus project from all subjects involved in the study, as well as the signing models shown in the pictures.

Data Availability Statement: Some of the data used in this study are publicly available at <https://doi.org/10.25592/dgs.corpus-3.0>. The full dataset can be made available upon request.

Acknowledgments: We would like to thank the DGS corpus participants and the DGS-Korpus annotation team for their contributions to the dataset we used for this study. Furthermore, we would like to thank Anne Wienholz for her advice on the quantitative data analysis. We thank the attendees of our presentation at the Grammar and Corpora conference 2022 in Ghent for their comments that helped us in the further development of this study. We would like to thank our reviewers and editors for their valuable feedback and insights.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A



Figure A1. Minimal pairs for the parameters' handshape, orientation, location, and movement.

Appendix B

| | | | |
|---|---|---|--|
|  <p>Handshape: 1d "one"</p> |  <p>Handshape: 1 "one"</p> |  <p>Handshape: 2d "two"</p> |  <p>Handshape: 2 "two"</p> |
|  <p>Handshape: 3d "three"</p> |  <p>Handshape: 4 "four"</p> |  <p>Handshape: 4d "four"</p> |  <p>Handshape: 5 "five"</p> |
|  <p>Handshape: 6d "six"</p> |  <p>Handshape: 7d "seven"</p> |  <p>Handshape: 8d "eight"</p> |  <p>Handshape: 9 "nine"</p> |
|  <p>handshape: 9d "nine"</p> |  <p>handshape: 10 "ten"</p> | | |

Figure A2. Numeral handshapes from 1 to 10, including variants.

Appendix C. DGS Examples with Glosses, Mouthings/Mouth Gestures, Translations, and Sources

| | | | | | | | |
|-----|--|--|----------------------|----------------------|--------------|-------------------------|--|
| (1) | AND2 | \$NUM-TEEN1:5 | CHILD2* | \$NUM-ONE-TO-TEN1A:5 | DEAF1A | \$NUM-ONE-TO-TEN1A:5 | |
| | | Fünfzehn | Kinder | Fünf | Taub | Fünf | |
| | HEAVY1B* | TO-HEAR2 | \$NUM-ONE-TO-TEN1A:5 | CI1* | \$GEST-OFF1^ | | |
| | | schwerhörig | fünf | | | | |
| | There were 15 kids: 5 were deaf, 5 were hard of hearing, and 5 had a CI. | | | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1245390 | | | Timecode: | 00:10:25:10-00:10:32:31 | |

| | | | | | | | |
|-----|--|--|--------|--------------|-----------------------|-------------------------|----------|
| (2) | MY1 | FREE1 | TIME5A | TO-BELIEVE2B | \$NUM-ONE-TO-TEN1A:6d | \$NUM-TENS1:2d | MEMBER4* |
| | | Freizeit | | Glaube | Sechszwanzig | | Mitglied |
| | I think the leisure [bowling] club has about 26 members. | | | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1583950 | | | Timecode: | 00:09:36:40-00:09:41:07 | |

| | | | | | | |
|-----|--|--|----------|-----------------------|-----------------------|-------------------------|
| (3) | \$INDEX1* | ALSO1A | \$INDEX1 | TO-TRY-OR-REHEARSAL2* | \$NUM-ONE-TO-TEN1A:6d | WEEK1A |
| | | | Aber | Probe | Sechs | Woche |
| | There was a probation period of six weeks. | | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1429964 | | | Timecode: | 00:10:22:37-00:10:24:34 |

| | | | | | | |
|-----|--|--|-------|-------------------------|--------------|-------------------------|
| (4) | I1 | TO-THINK1B | OLD8B | TO-THINK1B | BEGINNING1A* | TO-THINK1B |
| | | | | alt | anfang | |
| | \$NUM-ONE-TO-TEN1A:3d | OLD8B* | I1 | TO-KNOW-OR-KNOWLEDGE2B^ | NOTHING1B* | |
| | drei | alt | | [MG] | | |
| | I think I was three years old, but my memory doesn't go that far back. | | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1428225 | | | Timecode: | 00:00:01:49-00:00:07:16 |

| | | | | | |
|-----|---|--|------------|------------|-------------------------|
| (5) | MY1 | BOYFRIEND-GIRLFRIEND1 | BEEN1 | MONTH1* | |
| | Mein | Freund | Gewesen | Zwei monat | |
| | AUSTRALIA1* | \$INDEX1 | VACATION8B | BEEN1 | |
| | Australien | | Urlaub | Gewesen | |
| | My boyfriend spent two months on vacation in Australia. | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1289910 | | Timecode: | 00:08:25:07-00:08:28:08 |

| | | | | | |
|-----|---|--|--------|-----------|-------------------------|
| (6) | TO-COME1* | \$NUM-CLOCK1A:5 | UNTIL1 | EVENING2* | \$NUM-TEEN2B:1d |
| | mama | fünf uhr | bis | abend | elf uhr |
| | She [mom] came over around five in the afternoon and stayed until 11:00 at night. | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1176846 | | Timecode: | 00:08:39:16–00:08:24:42 |

| | | | | | | | |
|-----|---|--|-------|-----------|-------------------|-------------------------|--------|
| (7) | FREIBURG1* | LOCATION1B^ | FAR1* | FROM-TO3* | HOUR2B* | HALF6 | FAR1* |
| | freiburg | | [MG] | | drei stunden halb | | wei{t} |
| | Yet it [Freiburg] is also quite far away. It took us three [and a half] hours to get there. | | | | | | |
| | Public Corpus transcript: | https://doi.org/10.25592/dgs.corpus-3.0-text-1184756 | | | Timecode: | 00:11:28:48–00:11:33:08 | |

Appendix D

Table A1. List of type glosses and type entries in the Public DGS Corpus.

| Glossname | DOI of Superordinate Type |
|--------------------|--|
| WEEK1A | https://doi.org/10.25592/dgs.corpus-3.0-type-14967 |
| WEEK1B | https://doi.org/10.25592/dgs.corpus-3.0-type-48443 |
| WEEK1C | https://doi.org/10.25592/dgs.corpus-3.0-type-75108 |
| WEEK2 | https://doi.org/10.25592/dgs.corpus-3.0-type-16795 |
| YEAR1A | https://doi.org/10.25592/dgs.corpus-3.0-type-15127 |
| YEAR1B | https://doi.org/10.25592/dgs.corpus-3.0-type-15554 |
| \$NUM-CLOCK1A | https://doi.org/10.25592/dgs.corpus-2.0-type-72542 |
| CLOCK1 | https://doi.org/10.25592/dgs.corpus-3.0-type-16979 |
| \$NUM-ONE-TO-TEN1A | https://doi.org/10.25592/dgs.corpus-1.0-type-49944 |
| \$NUM-ONE-TO-TEN1B | https://doi.org/10.25592/dgs.corpus-1.0-type-50050 |
| \$NUM-ONE-TO-TEN1C | https://doi.org/10.25592/dgs.corpus-1.0-type-50052 |
| \$NUM-ONE-TO-TEN1D | https://doi.org/10.25592/dgs.corpus-1.0-type-50054 |
| \$NUM-TEEN1 | https://doi.org/10.25592/dgs.corpus-1.0-type-49945 |
| \$NUM-TENS1 | https://doi.org/10.25592/dgs.corpus-1.0-type-49948 |
| MONTH1 | https://doi.org/10.25592/dgs.corpus-3.0-type-15420 |
| DAY1A | https://doi.org/10.25592/dgs.corpus-3.0-type-14038 |
| DAY1B | https://doi.org/10.25592/dgs.corpus-3.0-type-13269 |
| DAY2 | https://doi.org/10.25592/dgs.corpus-3.0-type-18466 |
| HOUR2A | https://doi.org/10.25592/dgs.corpus-3.0-type-85194 |
| HOUR2B | https://doi.org/10.25592/dgs.corpus-3.0-type-54108 |
| HOUR2C | https://doi.org/10.25592/dgs.corpus-3.0-type-85822 |
| EURO1 | https://doi.org/10.25592/dgs.corpus-3.0-type-17418 |

Table A1. Cont.

| Glossname | DOI of Superordinate Type |
|--------------------|---|
| OLD5A | https://doi.org/10.25592/dgs.corpus-3.0-type-14440 |
| OLD8B | https://doi.org/10.25592/dgs.corpus-3.0-type-14503 |
| \$NUM-HUNDREDS1 | https://doi.org/10.25592/dgs.corpus-3.0-type-49953 |
| \$NUM-THOUSANDS1 | https://doi.org/10.25592/dgs.corpus-1.0-type-49954 |
| BEAUTIFUL1A | https://doi.org/10.25592/dgs.corpus-3.0-type-15596 |
| PLEASE1B | https://doi.org/10.25592/dgs.corpus-3.0-type-78494 |
| TO-COMPREHEND1 | https://doi.org/10.25592/dgs.corpus-3.0-type-15301 |
| TO-BE-SILENT1A | https://doi.org/10.25592/dgs.corpus-3.0-type-15047 |
| TO-SAW1 | https://doi.org/10.25592/dgs.corpus-3.0-type-13139 |
| CLASS1 | https://doi.org/10.25592/dgs.corpus-3.0-type-15218 |
| MILK2C | https://doi.org/10.25592/dgs.corpus-3.0-type-17832 |
| DIFFICULT1 | https://doi.org/10.25592/dgs.corpus-3.0-type-13152 |
| GERMAN-MARK | https://doi.org/10.25592/dgs.corpus-3.0-type-17377 |
| \$NUM-GERMAN-MARK1 | https://doi.org/10.25592/dgs.corpus-1.0-type-17919 |

Appendix E

Table A2. Absolute sign frequencies per age group.

| | 18–30 | 31–45 | 46–60 | 61+ | Totals |
|------------|-------|-------|-------|-----|--------|
| THOUSANDS1 | 110 | 133 | 133 | 86 | 462 |
| HUNDREDS1 | 106 | 165 | 216 | 205 | 692 |
| YEAR1 | 121 | 187 | 223 | 202 | 733 |
| MONTH1 | 43 | 64 | 42 | 36 | 185 |
| WEEK1 | 62 | 78 | 55 | 25 | 220 |
| DAY2 | 0 | 4 | 5 | 6 | 15 |
| HOUR2 | 118 | 155 | 149 | 112 | 534 |
| OLD8B | 1 | 24 | 16 | 22 | 63 |
| EURO1 | 26 | 37 | 20 | 5 | 88 |
| Totals | 587 | 847 | 859 | 699 | 2992 |

Notes

- Depending on the sign, it is possible to form 11 to 19 with a productive movement. However, in general, 11 and 12 have a tendency to be expressed by lexicalized suppletive forms in DGS.
- Pfau and Steinbach (2021) claim that in DGS only the numerals 1 to 5 can be incorporated in lexical base signs due to the two-handedness of the numerals from 6 to 10. In our corpus data, the two-handed numerals between 6 to 10 are incorporated. The movement of the lexical base sign is produced by both hands in this case.
- It is common in literature on sign languages to refer to specific sign types via labels in capital letters (called glosses). See Section 3.2. for more detail on glosses and the glossing conventions adhered to in the annotation of the DGS corpus. For this introduction, we use a simplified gloss annotation.
- Numeral and base sign share the same location in this instance, and the origin of the location in the affixed form is thus opaque. However, analyses of other incorporations show that the location matches the base sign.
- The handshape allomorphs for “one” in ASL as well as in DGS are unmarked. It is the selected thumb or index finger as in Figure 2b.
- <https://doi.org/10.25592/dgs.corpus-3.0>

- 7 See Schulder and Hanke (2022) for details on the FAIR and CARE policy within the project. Research on the total of the reference corpus requires an additional license agreement for researchers outside of the DGS-Korpus project. As all authors of this paper are researchers working in the project, we had unrestricted access to the data.
- 8 See Konrad et al. (2022, p. 7) for glossing conventions and the definition of the term gloss: “A gloss is a German word that corresponds—on the subtype level—to a core meaning (keyword) of the sign.”
- 9 Mouthing refers to the visible mouth movement of silently articulated German words. Mouth gestures are meaningful mouth movements not related to the articulation of a word and are annotated with the label [MG], subsuming all types of mouth gestures.
- 10 Deviant in this context is used as a label for tokens whose form in actual execution of the sign differs in some aspect to the form assigned to the type as citation form in iLex. It covers aspects of adaptations due to performance and signed context, such as assimilations, as well as aspects of variation and modification that have not yet been categorized in more detail. The annotation as “deviant” only indicates some difference to the citation form.
- 11 HamNoSys, short for Hamburg Notation System for Sign Languages, is a notation system for sign languages (Hanke 2004).
- 12 For more information on the prefixes used in the glosses, see Konrad et al. (2022, pp. 14–16).
- 13 In the DGS corpus, signs of the category \$NUM end by specifying the numeral handshape. This is accomplished by adding a qualifier that includes the numbers 1–10.
- 14 The qualifier 'numinc was also included in this study. It occurs only in addition to the qualifier 'q. As the differences between 'q and 'numinc + 'q are minor and irrelevant to our study, they will not be described here.
- 15 See Loos and Konrad (2022, p. 20) for a list of all qualifiers, their occurrences, functions, and codes as used in detailed corpus annotation.
- 16 The sign EURO1 is a relatively new sign as the Euro was introduced to Germany as a currency in 2002. The sign is produced with the handshape of extended index and middle finger, resembling the two lines of the € symbol. One of the signs used for the previous currency, D-Mark, has no fully specified base form and incorporates numerals from 1 to 10. It is possible that EURO1 started incorporating numerals due to this use of \$NUM–GERMAN–MARK1.
- 17 The asterisks are commonly used to indicate different levels of significance. They are used as follows: *** for $p < 0.001$, ** for $0.001 \leq p < 0.01$, and * for $0.01 \leq p < 0.05$.

References

- Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand. 1991. The Apparent Time Construct. *Language Variation and Change* 3: 241–64. [CrossRef]
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using Lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Brentari, Diane. 2012. 3. Phonology. In *Sign Language*. Edited by Roland Pfau, Markus Steinbach and Bencie Woll. Berlin and Boston: De Gruyter Mouton, pp. 21–54. [CrossRef]
- Bybee, Joan L. 2006. From Usage to Grammar: The Mind’s Response to Repetition. *Language* 82: 711–33. [CrossRef]
- Bybee, Joan L. 2011. Chapter 6. Usage-based Theory and Grammaticalization. In *The Oxford Handbook of Grammaticalization*, Online ed. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press, pp. 69–78. [CrossRef]
- Chinchor, Nancy. 1982. *Morphological Theory and Numeral Incorporation in American Sign Language*. Ann Arbor: Brown University.
- Dachkovsky, Svetlana. 2022. Emergence of a Subordinate Construction in a Sign Language: Intonation Ploughs the Field for Morphosyntax. *Glossa: A Journal of General Linguistics* 7: 1–41. [CrossRef]
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks: Sage.
- Frishberg, Nancy. 1975. Arbitrariness and Iconicity: Historical Change in American Sign Language. *Language* 51: 696. [CrossRef]
- Hanke, Thomas. 2004. HamNoSys—Representing Sign Language Data in Language Resources and Language Processing Contexts. Paper presented at the LREC2004 Workshop on the Representation and Processing of Sign Languages: From SignWriting to Image Processing. Information Techniques and Their Implications for Teaching, Documentation and Communication, Lisbon, Portugal, May 30; Edited by Oliver Streiter and Chiara Vettori. Paris: ELRA, pp. 1–6. Available online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/04001.pdf> (accessed on 12 May 2023).
- Hanke, Thomas, and Jakob Storz. 2008. iLex—A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. Paper presented at the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora, Marrakech, Morocco, June 1; Edited by Onno Crasborn, Eleni Efthimiou, Thomas Hanke, Ernst D. Thoutenhoofd and Inge Zwisserlood. Marrakech: European Language Resources Association (ELRA), pp. 64–67. Available online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/08011.pdf> (accessed on 13 February 2023).

- Hanke, Thomas, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in Size and Depth. Paper presented at the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, Marseille, France, May 16; Paris: European Language Resources Association (ELRA), pp. 75–82. Available online: <https://aclanthology.org/2020.signlang-1.12.pdf> (accessed on 13 February 2023).
- Hanke, Thomas, Reiner Konrad, and Gabriele Langer. 2023. Chapter 7. Exploring regional variation in the DGS Corpus. In *Advances in Sign Language Linguistics*. Edited by Ella Wehrmeyer. Amsterdam and Philadelphia: John Benjamins, pp. 192–218. [CrossRef]
- Hanke, Thomas, Reiner Konrad, Gabriele Langer, Anke Müller, and Sabrina Wähl. 2017. Detecting Regional and Age Variation in a Growing Corpus of DGS. Paper presented at the Workshop Corpus-Based Approaches to Sign Language Linguistics: Into the Second Decade, Birmingham, UK, July 24.
- Heine, Bernd, and Tania Kuteva. 2007. *The Genesis of Grammar: A Reconstruction*. Oxford and New York: Oxford University Press.
- Himmelman, Nikolaus P. 2004. Lexicalisation and Grammaticalization: Opposite or Orthogonal? In *What Makes Grammaticalization? A Look from Its Fringes and Its Components*. Edited by Walter Bisang, Nikolaus P. Himmelman and Björn Wiemer. Berlin and New York: Mouton de Gruyter, pp. 21–44. [CrossRef]
- Hopper, Paul J., and Elizabeth C. Traugott. 2003. *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press. [CrossRef]
- Isard, Amy, and Reiner Konrad. 2022. MY DGS—ANNIS: ANNIS and the Public DGS Corpus. Paper presented at the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources, Marseille, France, June 25; Edited by Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch and Marc Schulder. Marseille: European Language Resources Association (ELRA), pp. 73–79. Available online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/22034.pdf> (accessed on 13 February 2023).
- Janzen, Terry. 2012. 34. Lexicalization and Grammaticalization. In *Sign Language. An International Handbook*. Edited by Roland Pfau, Markus Steinbach and Bencie Woll. Berlin and Boston: De Gruyter Mouton, pp. 816–40. [CrossRef]
- Jones, Vanessa L. 2013. Numeral Incorporation in American Sign Language. In *Theses and Dissertations*. Grand Forks: The University of North Dakota, vol. 1551, Available online: <https://commons.und.edu/theses/1551/> (accessed on 11 May 2023).
- Konrad, Reiner, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. *Public DGS Corpus: Annotation Conventions/Öffentliches DGS-Korpus: Annotationskonventionen. Working Paper*. Version 4.1. Hamburg: DGS-Korpus-Projekt, June 14. [CrossRef]
- Ktejik, Mish. 2013. Numeral Incorporation in Japanese Sign Language. *Sign Language Studies* 13: 186–210. [CrossRef]
- Labov, William. 1963. The Social Motivation of a Sound Change. *WORD* 19: 273–309. [CrossRef]
- Lehmann, Christian. 1985. Grammaticalization: Synchronic Variation and Diachronic Change. *Lingua e Stile* 20: 303–18.
- Liddell, Scott K. 1996. Number Incorporating Roots & Non-Incorporating Prefixes in American Sign Languages. *Sign Language Studies* 92: 201–26.
- Liddell, Scott K., and Robert E. Johnson. 1986. American Sign Language Compound Formation Processes, Lexicalization, and Phonological Remnants. *Natural Language and Linguistic Theory* 4: 445–513. [CrossRef]
- Loos, Cornelia, and Reiner Konrad. 2022. *Detailed Annotation and Qualifiers. Working Paper*. Version 1. Hamburg: DGS-Korpus-Projekt, April 4. [CrossRef]
- Mathur, Gaurav, and Christian Rathmann. 2011. Two Types of Nonconcatenative Morphology in Signed Languages. In *Deaf around the World*, Online ed. Edited by Gaurav Mathur and Donna Jo Napoli. Oxford: Oxford University Press, pp. 54–82. [CrossRef]
- Meir, Irit. 2012. 5. Word Classes and Word Formation. In *Sign Language*. Edited by Roland Pfau, Markus Steinbach and Bencie Woll. Berlin and Boston: De Gruyter Mouton, pp. 77–111. [CrossRef]
- Pfau, Roland, and Markus Steinbach. 2011. Grammaticalization in Sign Languages. In *The Oxford Handbook of Grammaticalization*, Online ed. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press, pp. 683–95. [CrossRef]
- Pfau, Roland, and Markus Steinbach. 2021. Number in Sign Languages. In *The Oxford Handbook of Grammatical Number*. Edited by Patricia Cabredo Hofherr and Jenny Doetjes. Oxford: Oxford University Press, pp. 644–60. [CrossRef]
- Quer, Josep, Carlo Cecchetto, Caterina Donati, Carlo Geraci, Meltem Keleşir, Roland Pfau, and Markus Steinbach, eds. 2017. Part 2: Phonology. In *SignGram Blueprint. A Guide to Sign Language Grammar Writing*. Berlin and Boston: De Gruyter Mouton, pp. 17–68. [CrossRef]
- Sagara, Keiko, and Ulrike Zeshan. 2016. Semantic Fields in Sign Languages—A Comparative Typological Study. In *Semantic Fields in Sign Languages: Colour, Kinship and Quantification*. Edited by Ulrike Zeshan and Keiko Sagara. Berlin and Boston: De Gruyter Mouton, pp. 3–38. [CrossRef]
- Schulder, Marc, and Thomas Hanke. 2022. How to Be FAIR When You CARE: The DGS Corpus as a Case Study of Open Science Resources for Minority Languages. Paper presented at the 13th International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, June 20–25; Edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani and et al. Paris: European Language Resources Association (ELRA), pp. 164–73. Available online: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.18> (accessed on 13 February 2023).
- Schulder, Marc, Dolly Blanck, Thomas Hanke, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Susanne König, Reiner Konrad, Gabriele Langer, and et al. 2021. *Data Statement for the Public DGS Corpus. Working Paper*. Version 2. Hamburg: DGS-Korpus-Projekt, November 11. [CrossRef]

- Semushina, Nina, and Rachel I. Mayberry. 2019. Numeral Incorporation in Russian Sign Language: Phonological Constraints on Simultaneous Morphology. *Sign Language Studies* 20: 83–131. [CrossRef] [PubMed]
- Taub, Sarah F. 2001. *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge and New York: Cambridge University Press.
- Traugott, Elizabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press. [CrossRef]
- von Mengden, Ferdinand. 2010. Chapter I. Linguistic Numeral Systems. In *Cardinal Numerals: Old English from a Cross-Linguistic Perspective*. Berlin and New York: De Gruyter Mouton, pp. 12–71. [CrossRef]
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, and et al. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4: 1686. [CrossRef]
- Zeshan, Ulrike, Cesar Ernesto Escobedo Delgado, Hasan Dikyuva, Sibaji Panda, and Connie de Vos. 2013. Cardinal Numerals in Rural Sign Languages: Approaching Cross-Modal Typology. *Linguistic Typology* 17: 357–96. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Combinatorial Productivity of Spanish Verbal Periphrases as an Indicator of Their Degree of Grammaticalization

Mar Garachana ^{1,*} and María Sol Sansiñena ^{2,*}

¹ Departamento de Filología Hispánica, Teoría de la Literatura y Comunicación, Universitat de Barcelona, 08007 Barcelona, Spain

² Department of Linguistics, University of Leuven, 3000 Leuven, Belgium

* Correspondence: margarachana@ub.edu (M.G.); mariasol.sansiñena@kuleuven.be (M.S.S.)

Abstract: Studies on the constitution of the Spanish periphrastic system show that there is a great ease with which verbal periphrases admit different lexical items in the second verb slot as they go through their grammaticalization process. However, it has not been sufficiently explored whether the evolution of combinatorial patterns in near-synonymous periphrases follows similar grammaticalization paths. Adopting a constructionist, usage-based approach, we investigate the evolution of the so-called near-synonymous periphrases *dejar de* + INF and *parar de* + INF, as in *Deja de/Para de gritar*, ‘Stop shouting.’ More specifically, we discuss the semantic areas they cover, the functional distribution between the two throughout time, their evolution in terms of collostructional patterns, and their realized and potential productivity, paying special attention to the *Aktionsart* of the predicates in the Vinf slot. All tokens in the corpus were extracted from CORDE and analyzed in terms of morphosyntactic and semantic-pragmatic parameters, as well as contextual elements. We conduct a distinctive collexeme analysis to investigate which lexemes are strongly attracted or repelled by the Vinf slot in each construction. This analysis shows that the evolution of *parar de* + INF is not parallel to that of *dejar de* + INF and that there is a clear distribution of labor between the two constructions.

Keywords: Spanish verbal periphrases; productivity; grammaticalization; distinctive collexeme analysis

Citation: Garachana, Mar, and María Sol Sansiñena. 2023. Combinatorial Productivity of Spanish Verbal Periphrases as an Indicator of Their Degree of Grammaticalization.

Languages 8: 187. <https://doi.org/10.3390/languages8030187>

Academic Editor: Jeanine Treffers-Daller

Received: 3 March 2023

Revised: 4 July 2023

Accepted: 25 July 2023

Published: 7 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The objective of this work is to prove that the differences that allow us to distinguish so-called near-synonymous periphrases are related, at least partially, with the predicate types that appear in the position of the verb in the non-finite form. This is particularly true about phasal periphrases, whose meaning determines the *Aktionsart* of the verb in the non-finite form. To test our hypothesis, we will address the evolutionary study of two aspectual periphrases, namely, *dejar de* + INF and *parar de* + INF, as in (1a–1b):

- (1) a. *Por fin dej-ó de llov-er*
for end leave-IND.PST.3SG of rain-INF
‘At last, it stopped raining’
- b. *Cuando par-e de llov-er, tend-er-é la ropa*
when stop-SUBJ.PST.3SG of rain-INF hang.out-FUT-1SG the clothes
‘When it stops raining, I’ll hang out the clothes’

The periphrasis in (1a) has been documented in Spanish since the first written manifestations of the language. The periphrasis in (1b), on the other hand, is only found in texts from the 16th century onward. The existence of *dejar de* + INF, a consolidated verbal construction in the language since the medieval period, raises the question of how the most recent periphrastic construction arose and how it was integrated into the paradigm of phasal periphrases. More specifically, we wonder whether *parar de* + INF is born as a syntactic copy of *dejar de* + INF and in its expansion in the language is thus subordinated to the values of *dejar de* + INF, or whether it follows a free model, which would confirm the constructionist hypothesis that syntactic evolution affects not only saturated constructions

but also more abstract schemes (Hoffmann 2014, p. 161). Likewise, we will look into the changes in productivity of both periphrases throughout their history and what the implications of such changes are. Note that we use the concept of productivity in the sense of Barðdal (2008), that is, the capacity of a construction to gradually attract new lexical items.

To answer these questions, we will address the historical study of both periphrases from a corpus-based and collostructional perspective, which will allow us to measure the similarities and differences between the periphrases under study by investigating which lexemes are strongly attracted or repelled by the non-finite verb form (V_{inf}, in this case) slot in each construction. If the predicate types that appear as infinitives in the periphrasis *parar de* + INF replicate those that appear in *dejar de* + INF, we can argue that the paradigmatic attraction exerted by *dejar de* + INF on *parar de* + INF affects not only the meaning of the construction but also its combinatorics. In such a case, the analogic pressure exerted by *dejar de* + INF would affect the more saturated levels of the construction with *parar*. On the other hand, if differences relative to the predicate types are detected, this might suggest that the paradigmatic relations between near-synonymous periphrases suppose a distribution of values already from the first uses of the constructions. At this point, we want to point out that we use the term *analogy* as it is conceptualized in Barðdal (2008, pp. 2–3), that is, as “analogical extensions, i.e., extensions of a grammatical pattern based on only one model item, because of structural or semantic similarities between the two items.”

Before continuing, it is worth mentioning that the periphrastic character of *dejar de* + INF has not raised doubts and that it has been accepted as a member of the category of verbal periphrases by all the authors who have dealt with it (Yllera 1980; Gómez Torrego 1988, 1999; Gómez Manzano 1992; Olbertz 1998; Camus 2006a; RAE/ASALE 2009). On the other hand, the unanimity is not total with regards to the status of *parar de* + INF, for which links continue to be traced back to the lexical verb *parar*, ‘to stop.’ Thus, while some authors do not hesitate to consider that *parar de* + INF is a periphrasis (Gómez Torrego 1999; Camus 2006b; RAE/ASALE 2009; Aparicio Mera 2016), others argue that it only works as a periphrasis in contexts of negative modality (Fernández de Castro 1999, p. 267), where it expresses a continuative meaning with no traces of the sense of ‘stopping a movement’ (e.g., *No para de molestar* means ‘(s)he bothers all the time’). Finally, other authors exclude *parar de* + INF from the category of periphrases, regard it as a semi-auxiliary construction (Olbertz 1998, p. 114), or do not even include it in their studies (Gómez Manzano 1992).

In this paper, we are not going to discuss whether *parar de* + INF meets all the requirements that are considered defining of the paradigm of periphrases, since we start from a radial conception of grammatical categories, with some prototypical members (such as *tener que* + INF, ‘to have to + inf’) and others that may not meet all the defining criteria (such as *parar de* + INF). We follow the approach set out in previous studies by Garachana (2017) and Rosemeyer and Garachana (2019) and argue that *parar de* + INF is a grammatical construction in the sense of construction grammar, i.e., it is a conventionalized pairing of form and meaning. Given that in practically all the contexts of use of *parar de* + INF, it is feasible to replace it with *dejar de* + INF, and given that in the cases in which said alternation is not possible, this is due to pragmatic or usage issues, we fail to see any reason for not contemplating *parar de* + INF as a non-prototypical periphrastic construction. However, we must highlight the semantic differences between the lexical meaning of *parar* and its employment as an auxiliary verb. Indeed, the lexical meaning of *parar* refers to the interruption of a movement; thus, *parar el coche* (lit. ‘to stop the car’) means ‘interrupting the movement of a car.’ On the contrary, the interruption that is predicated in the periphrasis, especially with certain verbs, does not refer to any movement but must be interpreted in terms of aspect. In this sense, *ha parado de llover*, ‘it has stopped raining’ does not imply the cessation of any movement but rather that the event expressed by the infinitive *llover*, ‘to rain’ is interrupted. That is, from the moment that an infinitive appears in the position of complement of the preposition, *parar* is reanalyzed as an auxiliary, and ensemble comes to mean interruptive aspect. Considering that *parar* in *parar de llover* ‘to stop raining,’ or even in *parar de molestar* ‘to stop bothering’ or *parar de beber* ‘to stop drinking,’ is equivalent to

stopping a movement or activity, it implies confusing a semantic interpretation with an aspectual one: the construction *parar de + INF* is used to explain how the event designated by the infinitive develops. Certainly, a metaphorical relationship can be perceived between *parar* ‘to stop’ and *parar de + INF* ‘cessative egressive aspect,’ however, an absolute identification between the lexical verb and the auxiliary verb in the periphrasis is not feasible; only the *persistence* (in the sense of Hopper 1991) can be identified (for a more detailed discussion on this issue, cf. Garachana 2021).

The rest of the article is organized as follows: Section 2 briefly discusses the theoretical framework, corpus data, and methods used. Section 3 presents a general description of the semantics of “*dejar de + INF*” and “*parar de + INF*” in contemporary Spanish. Section 4 discusses the evolution of both periphrases under study throughout time. In this section, we also look at the realized and potential productivity of both periphrases, paying special attention to the *Aktionsart* of the predicates. Section 5 discusses the results of our qualitative and quantitative data analysis, and Section 6 summarizes our findings and advances some final conclusions.

2. Theoretical and Methodological Framework

We investigate the evolution of the near-synonymous Spanish periphrases *dejar de + INF* and *parar de + INF*, adopting a constructionist, usage-based approach. Specifically, our work is framed in the models of Construction Grammar (as developed by Langacker 1987, 2003; Fillmore 1996; Croft 2000; Croft and Cruise 2004; Goldberg 2006; Hoffmann 2013, a.o.) and Diachronic Construction Grammar (e.g., Fried 2009; Traugott and Trousdale 2013; Barðdal et al. 2015).

To carry out this research, all documented instances of both periphrases in the diachronic corpus of the Real Academia Española (2007) CORDE (*Corpus diacrónico del español*) were extracted. This written corpus constitutes the largest historical data bank in Spanish, whose ecdotic reliability can be verified from the work carried out by Rodríguez Molina and Octavio de Toledo y Huerta (2017). CORDE covers a historical period that spans from the origins of the language until 1974 and contains texts of different textual typologies. We obtained a total of $n = 308$ tokens of valid data for the study of *parar de + INF* and $n = 31,362$ tokens of *dejar de + INF*. The datasets on which we base our study come from the documentary archive of the GRADIA group.

Each token in our datasets was analyzed in terms of morphosyntactic and lexicogrammatical parameters, such as predicate type and lexical aspect of the predicate; semantic-pragmatic parameters, such as meaning of the construction and communicative function; and contextual elements, such as, e.g., sentence modality and adverbial complements with scope over *dejar de + INF* and *parar de + INF*. Next, we conducted a collostructional analysis to investigate which lexemes are strongly attracted or repelled by the non-finite verb form slot in the construction, both per period and throughout time. This was done through a distinctive collexeme analysis (Gries and Stefanowitsch 2004), which is often used to study one slot in two (or more) similar constructions and is known to be useful when applied to alternating pairs of constructions. We found distinctive collexeme analysis appropriate to be applied to the analysis of diachronic corpus data, as it is a very useful tool for the study of grammaticalizing constructions (Hilpert 2006). The collostructional analysis performed on both periphrastic constructions focused on their non-finite verb slot (Vinf). Hence, the first step in the analysis was semi-automatically identifying the verbs used as infinitives with *dejar de* and with *parar de*. For each different type, the number of tokens was counted, including the hapax legomena, i.e., the items that appear only once in the corpus. The statistical significance of the association between the construction and the lexical item in the Vinf slot was calculated by means of a Fisher-Yates exact test, and the results are presented in the form of the negative base-10 logarithm of the p-value. Finally, we also looked at the realized and potential productivity (see Baayen 2009) of both periphrases, paying special attention to the *Aktionsart* of the attracted or repelled predicates.

3. The Semantics of *dejar de* + INF and *parar de* + INF in Contemporary Spanish

The periphrases *dejar de* + INF and *parar de* + INF are often described as phasal periphrases (Yllera 1980; Olbertz 1998; Fernández de Castro 1999; Camus 2006a, 2006b; RAE/ASALE 2009). They both express interruptive and continuative meanings. The first ones are exemplified in (2): in (2a), dinner is not over but is interrupted; similarly, in (2b), the interlocutor is asked to interrupt the movement that is causing the speaker to get irritated.

- (2) a. *Dej-ó* *de* *cenar* *porque* *lo* *llam-ar-on*
 leave-IND.PST.3SG of dine-INF because 3SG.M.ACC call-IND.PST-3PL
 ‘He **stopped eating dinner** because someone called him’
- b. *¡Par-a* *de* *mov-er-te,* *por* *favor!*
 stop-IMP of move-INF-PRON for favour
 ‘**Stop moving**, please!’

The interruptive value of these constructions, when the non-finite verb is an accomplishment or an achievement, equates them to negative implicative verbs since the action, when interrupted, does not take place anymore; e.g., *dejar de cobrar un impuesto* (lit. ‘to stop charging a tax’) means that the tax is no longer being collected. In this way, when the periphrasis falls under the scope of a negative polarity term, an affirmative interpretation can be activated. Thus, the ensemble ‘negative + interruptive periphrasis’ comes to signify the continuation of the action expressed by the verb in the infinitive. In (3a), the periphrases *dejar/parar de* + INF indicate that a person continues reading a book. The interruptive and continuative meanings require verbs expressing processes, accomplishments, and non-permanent states in the Vinf slot. When achievements or permanent states appear in the Vinf slot, the interruptive and continuative meanings add a nuance of iteration. Thus, in (3b), Sara is asked to stop jumping once and again, and in (3c), Juan thanks his mother repeatedly.

- (3) a. *No* *dej-alpar-a* *de* *le-er* *ese* *libro*
 NEG leave/stop-IND.PRS.3SG of read-INF that book
 ‘(S)he **doesn’t stop reading** that book’
- b. *¡Sara,* *dej-alpar-a* *de* *salt-ar!*
 Sara leave/stop-IMP of jump-INF
 ‘Sara, **stop jumping!**’
- c. *Juan* *no* *par-ab-a/dej-ab-a* *de* *agradec-er* *a* *su* *madre*
 Juan NEG leave/stop-IND.PST-3SG of thank-INF to his mother
 ‘Juan **kept on thanking** his mother’

These interruptive and continuative values allow us to characterize the periphrasis *parar de* + INF. However, *dejar de* + INF is not so closely linked to the expression of phasal values (see Section 4). In fact, the continuative values have always had a very low frequency of use in the history of the periphrasis, and the more frequent interruptive values come into direct competition with other meanings in which the periphrasis seems to have specialized, namely the assertive ones (4) and those that express the negation of the event designated by the infinitive (5). In assertive meanings, *dejar de* + INF is used to formulate attenuated assertions, and commissive and directive speech acts in a polite manner. That is, in these cases, the periphrasis is always negated and has either a neutral assertive value (4a), expresses a vehement assertion (4b), or constitutes an attenuated or polite formulation of a request (4c) or a suggestion (4d).

- (4) a. *no dej-ó de advert-ir que ten-í-an* [...]

NEG leave-IND.PST.3SG of notice-INF COMP have-IND.PST-3PL
 ‘She **noticed** that they had [...]’
 (Alegría, *Perros hambrientos*, 1939, CORDE)
- b. *no dej-o de conoc-er que lo que pid-es es bueno*

NEG leave-IND.PRS.1SG of know-INF COMP 3SG.M.ACC COMP
 ask.for-IND.PRS.2SG be.IND.PRS.3SG good
 ‘I **know** that what you ask for is good’
 (Menéndez Pelayo, *Orígenes de la novela*, 1905, CORDE)
- c. *Adiós, no dej-e-n de avis-ar*

goodbyeNEG leave-SUBJ.PRS-2PL of notice-INF
 ‘Goodbye, **please let us know.**’
 (Ignacio Aldecoa, *El fulgor y la sangre*, 1954, CORDE)
- d. *No dej-e-s de ir a la Villa d’Este*

NEG leave-SUBJ.PRS-2SG of go-INF to the Villa d’Este
 ‘**Be sure to go to Villa d’Este.**’
 (Pedro Salinas, *Correspondencia*, 1951, CORDE)

Utterances in which *dejar de + INF* makes it possible to negate the event expressed by the infinitive are usually activated when the verb in the infinitive is an achievement or if the periphrasis is temporally delimited. In (5a), for instance, *había dejado de aparecer* is interpreted in the sense of ‘she no longer appeared.’ The sense of negation is also activated very often in periphrastic chains. In (5b), for example, *no podemos dejar de advertir* should be read as ‘we cannot not point out’ (for a detailed exposition of the values of *dejar de + INF*, cf. Garachana n.d.).

- (5) a. *hab-í-a dej-ado de aparec-er por sus habitaciones*

have-IND.PST.3SG leave-PTCP of appear-INF for their bedrooms
 ‘(S)he **had ceased to appear** in their rooms’
 (Martín Virgil, *Los curas comunistas*, 1968, CORDE)
- b. *no pod-emos dej-ar de advertir que* [...]

NEG can-IND.PST.3SG leave-INF of notice-INF COMP
 ‘**We cannot fail to point out** that [...]’
 (Malpica, *El desarrollismo en el Perú*, 1974, CORDE)

4. Qualitative and Quantitative Analysis of the Corpus Data

In this section, we will discuss the frequency of use and the realized and potential productivity of the two periphrases under analysis, both per period, i.e., independently of the other centuries, and throughout time, i.e., establishing a comparison among centuries. Then, we will present the results of the distinctive collexeme analysis of the data from CORDE, and we will interpret the results in relation to the *Aktionsart* of the lexemes that enter into the *Vinf* slot and that are significantly preferred by either one construction or the other. Finally, these outcomes will allow us to comment on the semantic evolution of *dejar de + INF* and *parar de + INF* in the history of Spanish.

4.1. Frequency of *dejar de + INF* and *parar de + INF* in CORDE

In addition to the semantic differences that we have discussed in the previous section, if we look at the absolute frequency of use of *dejar de + INF* and *parar de + INF* in the language throughout time, we observe significant divergences (see Table 1 below).

Table 1. Tokens of *dejar de + INF* y *parar de + INF* in CORDE per century.

| Construction | 13th | 14th | 15th | 16th | 17th | 18th | 19th | 20th |
|-----------------------|------|------|------|--------|------|------|------|------|
| <i>dejar_de + INF</i> | 349 | 219 | 1268 | 10,058 | 5363 | 2151 | 5290 | 6664 |
| <i>parar_de + INF</i> | 0 | 0 | 0 | 13 | 28 | 7 | 56 | 202 |

In this section, we will present a comparison of the frequency of use of both periphrases throughout the history of Spanish, normalizing the frequency scores to a common base

per million words (see Figure 1 below). The results show that, while *dejar de* + INF reaches very high frequencies of use throughout the entire history of Spanish, *parar de* + INF moves along with very low relative frequencies, which do not manage to exceed one appearance per million words until the 19th century.

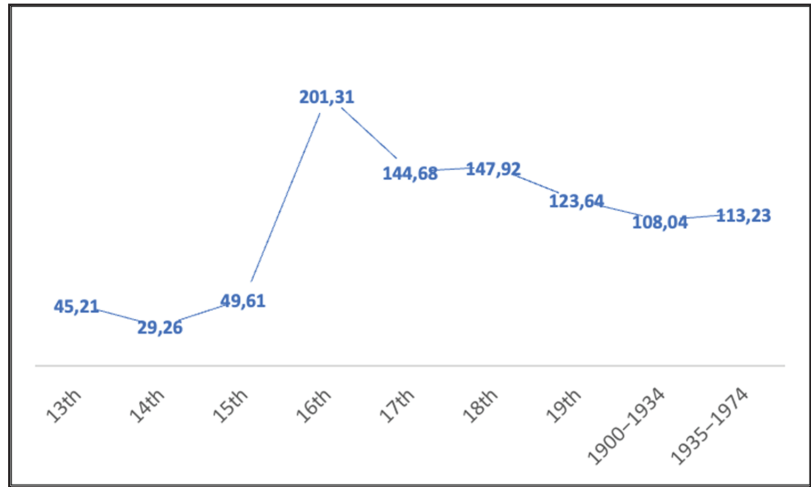


Figure 1. Normalized frequency of *dejar de* + INF per million words.

If we look at Figure 1 below, we can see that already in the medieval period, *dejar de* + INF was used with a frequency that oscillates between 25 tokens per million words in the century in which its presence is lowest and 55 tokens per million words in the 15th century, which marks the moment of maximum use of the periphrasis in that period. Its frequency reaches 201.51 tokens per million words in the 16th century, and the 17th and 18th centuries stand at about 140 appearances per million words each (140.82 in the 17th and 148.68 in the 18th century). At the end of the so-called “First Modern Spanish” (ca. 1675–1825), a slight decrease in the frequency of the periphrasis is detected, which possibly reflects the incipient establishment of *parar de* + INF in the language.

The high frequency documented in the 16th century must be properly contextualized since it is distorted by the predilection that some authors show for this structure in this period. This is not the place to properly assess the disproportionate increase in the normalized frequency of *dejar de* + INF in the 16th century, but it is significant that among the group of ‘anonymous authorships,’ we have documented the periphrasis 546 times only in the category ‘long stories.’ Works by a known author in which the periphrasis is well documented include, in addition to long narratives, historiographical texts (which in the 16th century were stylistically close to long narratives) and religious texts. These findings suggest that the high frequency of use of the periphrasis in the 16th century may be due to stylistic preferences. We observed that individual authors such as Santa Teresa de Jesús (435 tokens), Fray Luis de Granada (306 tokens), Antonio de Torquemada (310 tokens), and Gonzalo Fernández de Oviedo (558 tokens) tend to use *dejar de* + INF much more frequently than other authors in the same or other time periods.

It could be assumed that the high frequency of use of periphrasis in the works of these authors could be due to the fact that the CORDE contains a greater number of texts by these particular authors than by others. However, if we compare, for example, the use of *dejar de* + INF in two contemporary authors such as Santa Teresa de Jesús and Miguel de Cervantes, we observe that this high frequency of use of the periphrasis is not related to the length of the texts by these authors. Indeed, Santa Teresa de Jesús uses *dejar de* + INF on 435 occasions, and, in CORDE, the texts by this author make up a total of

560,405 words, which means a frequency of 776 tokens per million words. Instead, Miguel de Cervantes resorts to *dejar de* + INF on 45 occasions, and in CORDE, his works represent a total of 1,449,640 words, which means a frequency of use of 31 tokens per million words. These data seem to prove that the use of periphrasis obeys the rhetorical tendencies of the 16th century. In fact, as we will point out in the discussion section, this periphrasis was progressively more and more used within specific constructions, such as auxiliary chains. Furthermore, after the 16th century, this periphrasis tended to be used with non-aspectual meanings for the expression of pragmatic-discursive meanings, particularly for mainly assertive, commissive, and directive speech acts. It is not surprising, given the communicative relevance of these types of speech acts, that at one point in the history of Spanish, writers used this construction profusely for stylistic purposes.

The normalized frequency of *parar de* + INF presents a construction that is poorly consolidated in the language, at least in the written language recorded in CORDE. The low frequency of use of *parar de* + INF is reflected in Figure 2, adapted from Garachana (2021), which highlights the modern nature of the construction.

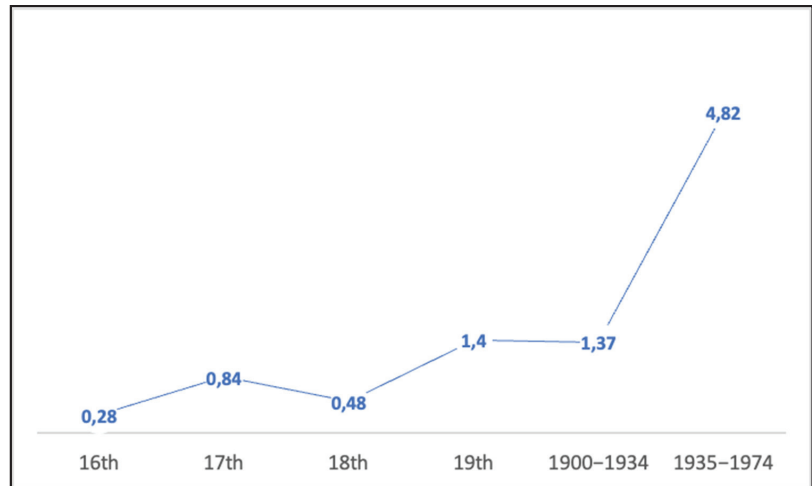


Figure 2. Frequency of *parar de* + INF per million words.

According to the data contained in Figure 2, the progressive expansion of *parar de* + INF in Spanish started in the 16th century. During the Spanish Golden Age (16th–17th c.) and the First Modern Spanish (ca. 1675–1825), the frequency of *parar de* + INF did not reach one token per million words. The 19th century shows a slight rise in the use of construction, which should not have been unrelated to the fact that, at this time, some features of orality are emerging in the written language, a process possibly favored by the rise of literary realism and naturalism, which exalted the reflection of everyday language in texts. However, the first moment in which a significant change in the frequency of use of the construction is observed is in the last quarter of the 20th century, when the normalized frequency of *parar de* + INF reaches four tokens per million words. In the second half of the 20th century, we documented frequencies between 4 and 7 occurrences of *parar de* + INF per million words.

4.2. Productivity of the Constructions *dejar de* + INF and *parar de* + INF

The comparative study carried out required considering the *syntactic productivity* of *dejar de* + INF and *parar de* + INF. Specifically, we measure the potential of both constructions to attract new verbal types to the Vinf slot, i.e., we measure the extensibility of both constructions. This is a question that has already been addressed in existing publications

on the history of the Spanish language (Rodríguez Molina 2004, 2007; Garachana 2016, 2022). In these works, it was pointed out that the key to the establishment of a periphrasis in the Spanish grammatical system lies in the gradual incorporation of new verbs into the verbal construction. We will resume this line of studies here and combine it with a focus on productivity in order to be able to properly quantify the extensibility of the verbal periphrases under study and their productivity at different moments in the history of Spanish.

If we look at the *realized productivity* of both periphrases, i.e., the number of word types of category C in a corpus of N tokens, we observe that *parar de + INF* is more productive than *dejar de + INF*, as Table 2 shows. According to Figures 1 and 2 above, *dejar de + INF* has a higher frequency of use than *parar de + INF*. However, that does not translate into greater productivity.

Table 2. Realized productivity of *dejar de + INF* and *parar de + INF* in the history of Spanish.

| Construction | 13th | 14th | 15th | 16th | 17th | 18th | 19th | 20th |
|-----------------------|------|------|------|------|------|------|------|------|
| <i>parar de + INF</i> | 0 | 0 | 0 | 0.92 | 0.85 | 1 | 0.48 | 0.37 |
| <i>dejar de + INF</i> | 0.23 | 0.29 | 0.23 | 0.07 | 0.11 | 0.22 | 0.13 | 0.12 |

Meanwhile, as Baayen (2009, p. 905) points out, there are aspects of productivity that are not well represented by a category’s realized productivity, as a high realized productivity “does not imply that its [...] potential productivity will be high as well.” Thus, we also calculated the *potential productivity* of both periphrases per century. The potential productivity of a rule “is estimated by its hapax legomena in the corpus divided by the total number of its tokens N(C) in the corpus: $P = V(1, C, N)/N(C)$ ” (Baayen 2009, p. 902) and expresses the growth rate of the lexemes of the category C itself. The results obtained, presented in Table 3 below, show that the potential productivity of the periphrases under study reveals an inversely proportional relationship with their normalized frequency. Furthermore, this finding supports the claim that the potential productivity is inversely proportional to the degree of grammaticalization of grammatical constructions (Bybee 2003; Bybee and Torres-Cacoullous 2009; Torres Cacoullous 2012; Aaron 2010; Copple 2011).

Table 3. Potential productivity of *dejar de + INF* and *parar de + INF* in the history of Spanish.

| Construction | 13th | 14th | 15th | 16th | 17th | 18th | 19th | 20th |
|-----------------------|------|------|------|------|------|------|------|------|
| <i>parar de + INF</i> | 0 | 0 | 0 | 0.84 | 0.75 | 1 | 0.39 | 0.22 |
| <i>dejar de + INF</i> | 0.12 | 0.19 | 0.12 | 0.03 | 0.04 | 0.12 | 0.06 | 0.05 |

The grammaticalization of a verbal periphrasis implies, among other changes, the progressive elimination of restrictions that affect verbs that may appear in the position of the non-finite verb form (vid. Garachana 2017). That is, the more grammaticalized a construction is, the higher the number of types that can function in such a position, so that the possibility that new verb forms appear in this position will eventually be increasingly limited. This is precisely what can be seen in Table 3: The potential productivity of *parar de + INF* is much higher than that of *dejar de + INF*. However, with the course of the centuries, we also observe that this potential productivity becomes gradually more and more limited, which is symptomatic of the progressive consolidation of *parar de + INF* in the Spanish grammar. Even so, the potential productivity of *parar de + INF* was still higher than that of *dejar de + INF* in the first centuries in which this periphrasis was used in the language. From the perspective of productivity, these data, which highlight the weakest settlement and fixation of *parar de + INF* in the language, offer us valuable information to respond to the discussions that remain in grammatical studies in relation to deciding whether *parar de + INF* is a verbal periphrasis or a ‘semi-periphrastic construction.’ The restrictions that are detected in relation to the verb types that may appear as *Vinf* connect directly with the

potential productivity of the periphrasis. The grammatical fixation of *dejar de* + INF does not raise doubts since the lexical restrictions that affect the verbs that function as infinitives are very limited. However, that means that its potential productivity is lower since, in its expansion as verbal periphrasis in Spanish, it has already reached most of the verb types. *Parar de* + INF, on the other hand, still shows a greater capacity to admit new verb types precisely because it has advanced less in its grammaticalization process.

Besides, we also observed that, if we look at the tokens of both constructions together, out of the 1655 different types documented, only 65 of them are shared by the two constructions, which represents an extremely low proportion. This suggests that the two constructions present a very limited overlap in their Vinf slots, not only in terms of semantic fields but also in terms of specific lexemes. To disentangle this issue, we now move on to the structural study.

4.3. Distinctive Collexeme Analysis

The quantitative analysis of the CORDE data includes a distinctive collexeme analysis. In this subsection, we will present and discuss the results of such an analysis. We looked at the dataset from two perspectives: (i) ignoring diachrony, we were able to check which lexemes (Vinf) are attracted by either construction in general terms (see Section 4.3.1); and (ii) dividing the dataset into periods, we could check which lexemes are attracted by one or the other construction in each period, and, in turn, we could review the attraction of each of these lexemes in all periods (see Section 4.3.2). A methodological sidenote is in order at this point. The values provided are based on the code proposed by Levshina (2015) and are presented as a base-10 logarithm of the p-value of the Fisher-Yates exact test. The value is positive when the occurrences are higher than expected, i.e., there is a preference for *dejar de*, and negative when they are lower than expected, i.e., there is a preference for *parar de*. Absolute values higher than 1.3 are classified as instances of ‘categorical preference’ and they are indicated with blue lines on the graphs presented in this subsection. For example, a graph showing verbs with preferences in the 20th century shows in blue the trajectory through the centuries of those verbs with <-1.3 or >1.3 in the 20th century, regardless of their values in the other centuries. Let us now present the findings of the quantitative analysis from the two above-mentioned perspectives.

4.3.1. In General

Table 4 shows the n of lexemes (Vinf) that prefer one or the other construction (collstructional value ‘collStr’ is higher than 1.3 or lower than -1.3), or none, throughout the centuries. We can observe that most of the lexemes do not prefer either variant.

Table 4. n of lexemes that prefer one or the other construction, or none, in the CORDE.

| Preference | Quantity |
|-----------------|----------|
| <i>dejar de</i> | 5 |
| none | 1594 |
| <i>parar de</i> | 64 |

Next, we can observe the lexemes that prefer *dejar de* + INF and their collStr value in Table 5.

Table 5. Lexemes that prefer *dejar de* + INF and collStr value.

| Lexeme | collStr |
|------------------------|---------|
| <i>ser</i> ‘to be’ | 24,231 |
| <i>tener</i> ‘to have’ | 4342 |
| <i>hacer</i> ‘to do’ | 3514 |
| <i>haber</i> ‘to have’ | 1782 |
| <i>ver</i> ‘to see’ | 1609 |

It is interesting to notice that the lexemes that prefer *dejar de* + INF in the dataset are mostly states, contrary to what has been claimed in the literature (cf. Olbertz 1998; Camus 2006a). *Deja de ser tan (insistente)*, for instance, means ‘give up this habit or way of being (insistent).’ In this sense, we can argue that *dejar de* + INF is often used to characterize the entity, as in *No deja de ser paradójico*, which means *Es paradójico*, ‘It’s paradoxical’ or *No dejas de tener razón*, which means *Tienes razón*, ‘You’re right.’

Now, Table 6 presents the lexemes that prefer *parar de* + INF and their collostructional value. Note that, since the values are calculated on the basis of *dejar de*, we consider as lexemes that prefer *parar de* those that demonstrate a significant rejection for *dejar de*. Therefore, the values in Table 6 are negative.

Table 6. Lexemes that prefer *parar de* + INF and collStr value.

| Lexeme | collStr | Lexeme | collStr | Lexeme | collStr | Lexeme | collStr |
|----------------------------------|---------|---|---------|-------------------------------------|---------|------------------------------------|---------|
| <i>contar</i> ‘to tell’ | −45,805 | <i>quejarse</i> ‘to complaint’ | −3034 | <i>refregar</i> ‘to scrub’ | −2012 | <i>vocear</i> ‘to holler’ | −1713 |
| <i>remover</i> ‘to stir’ | −17,019 | <i>subir</i> ‘to go up’ | −2855 | <i>retroceder</i> ‘to move back’ | −2012 | <i>trabajar</i> ‘to work’ | −1649 |
| <i>llover</i> ‘to rain’ | −9892 | <i>echar</i> ‘to throw’ | −2336 | <i>roznar</i> ‘to bray’ | −2012 | <i>agitar</i> ‘to shake’ | −1539 |
| <i>hablar</i> ‘to speak’ | −8533 | <i>mover</i> ‘to move’ | −2207 | <i>trajinar</i> ‘to bustle’ | −2012 | <i>regañar</i> ‘to scold’ | −1539 |
| <i>hervir</i> ‘to boil’ | −8040 | <i>bailar</i> ‘to dance’ | −2131 | <i>trapear</i> ‘to mop’ | −2012 | <i>revolotear</i> ‘to flutter’ | −1539 |
| <i>revolver</i> ‘to stir’ | −6187 | <i>acumular</i> ‘to accumulate’ | −2012 | <i>tronar</i> ‘to thunder’ | −2012 | <i>roltar</i> ‘to let go’ | −1539 |
| <i>beber</i> ‘to drink’ | −4876 | <i>aporracear</i> ‘to beat’ | −2012 | <i>zaherir</i> ‘to taunt’ | −2012 | <i>trinar</i> ‘to trill’ | −1539 |
| <i>correr</i> ‘to run’ | −4301 | <i>chancearse</i> ‘to jest’ | −2012 | <i>clamar</i> ‘to cry out’ | −1986 | <i>vomitarse</i> ‘to throw up’ | −1539 |
| <i>coocer</i> ‘to cook’ | −4135 | <i>comerciar</i> ‘to trade’ | −2012 | <i>cantar</i> ‘to sing’ | −1933 | <i>cegar</i> ‘to blind’ | −1416 |
| <i>bullir</i> ‘to boil’ | −3551 | <i>contrastar</i> ‘to contrast’ | −2012 | <i>derrocar</i> ‘to overthrow’ | −1713 | <i>exclamar</i> ‘to exclaim’ | −1416 |
| <i>provocar</i> ‘to provoke’ | −3551 | <i>dejar</i> ‘to leave’ | −2012 | <i>frotar</i> ‘to rub’ | −1713 | <i>tragar</i> ‘to swallow’ | −1416 |
| <i>rezongar</i> ‘to grumble’ | −3551 | <i>disparatar</i> ‘to talk nonsense’ | −2012 | <i>insultar</i> ‘to insult’ | −1713 | <i>reír</i> ‘to laugh’ | −1405 |
| <i>nevar</i> ‘to snow’ | −3253 | <i>llevar</i> ‘to carry’ | −2012 | <i>lisonjear</i> ‘to flatter’ | −1713 | <i>dar</i> ‘to give’ | −1404 |
| <i>gritar</i> ‘to shout’ | −3252 | <i>parlotear</i> ‘to chatter’ | −2012 | <i>mecer</i> ‘to rock’ | −1713 | <i>acrecentar</i> ‘to increase’ | −1322 |
| <i>tocar</i> ‘to touch’ | −3110 | <i>rebullir</i> ‘to start moving’ | −2012 | <i>preparar</i> ‘to prepare’ | −1713 | <i>fregar</i> ‘to scrub’ | −1322 |
| <i>protestar</i> ‘to protest’ | −3034 | <i>rechinar</i> ‘to grind’ | −2012 | <i>retorcer</i> ‘to twist’ | −1713 | <i>toser</i> ‘to cough’ | −1322 |

With *parar de*, there are 145 types documented in total, and 64 of them show a significant attraction to the construction. *Contar*, ‘to tell,’ the lexeme most attracted to *parar de*, appears with *dejar de* in only 324 instances, which are few if you consider the total number of tokens of *dejar de* + INF in the corpus. Among the 64 types that show a significant attraction to *parar de*, we find communication verbs such as *tell*, *speak*, *provoke*, *grumble*, and *yell*; predicates related to cooking, such as *boil* and *cook*; meteorological verbs, such as *rain* and *snow*; and physical contact verbs, such as *remove*, *stir*, and *touch*. Some of these are processes and do not imply completion or culmination. Many of these lexemes denote an activity or movement -they are not stative- and they generally take an animate subject.

4.3.2. Per Period

We applied the same procedure to samples from specific periods. In order to do so, we grouped together the data until the 16th century because before that century, there were no documented examples of the periphrasis *parar de* + INF in the CORDE. We also grouped together the 17th and 18th centuries to get a more balanced normalized frequency per period. The nineteenth century already marked a noticeable increase in the frequency of use of *parar de* + INF, which, in turn, increased considerably in the 20th century. These groupings are motivated by the behavior of the periphrasis *parar de* + INF in the corpus rather than by external factors. The n of lexemes with preferences for either variant or none of them in each period is shown in Table 7. Throughout time, both periphrases seem to ritualize certain types. And with a higher degree of grammaticalization, each seems to prefer more predicates over time.

Table 7. Lexemes that prefer *parar de* + INF, *dejar de* + INF or none per period.

| Period | None | <i>parar de</i> | <i>dejar de</i> |
|------------|------|-----------------|-----------------|
| until–16th | 845 | 7 | 0 |
| 17th–18th | 809 | 16 | 1 |
| 19th | 716 | 18 | 1 |
| 20th | 808 | 35 | 3 |

As can be observed in Table 7, no lexeme shows a significant preference for *dejar de* before the 17th century. From this century on, *ser* ‘to be’ becomes the lexeme most closely linked to *dejar de* + INF. We also see that *tener* ‘to have’ and *existir* ‘to exist’ are only documented with *dejar de* in the 20th century. These three lexemes are very frequent with *dejar de*, and they never occur with *parar de* (see Tables 8 and 9).

Table 8. Lexemes that prefer *dejar de* per period.

| Lexeme | collStr | Frequency | Period |
|---------------------------|---------|-----------|-----------|
| <i>ser</i> ‘to be’ | 2375 | 1158 | 17th–18th |
| | 6283 | 1247 | 19th |
| | 24,910 | 1682 | 20th |
| <i>tener</i> ‘to have’ | 3414 | 276 | 20th |
| <i>existir</i> ‘to exist’ | 1821 | 156 | 20th |

In the same way, we can list the verbs that prefer *parar de*, which are, in general, infrequent in the Vinf position of either construction. If one looks at the absolute frequency in the corpus, they tend to occur more in the context of *dejar de*, just because the construction itself is much more frequent. Since there are 68 lexemes that prefer *parar de* over *dejar de* in at least one period, we will present the data in parts. First, Table 9 shows the lexemes that only occur in the construction *parar de* + INF in a certain period and are not documented with *dejar de* + INF. Except for four lexemes in the 20th century subcorpus, which occur twice with *parar de*, the others are only documented once. Throughout time, we have observed a

gradual increase in the use of communication verbs with negative connotations, which are marked in bold here.

Table 9. Lexemes that are only documented with *parar de* in particular periods.

| Period | Lexeme | Frequency | collStr |
|------------|---|-----------|---------|
| until–16th | <i>bullir</i> ‘to boil’, <i>levar</i> ‘to carry’, <i>lisonjear</i> ‘to flatter’ | 1 | –2900 |
| 17th–18th | <i>aporracear</i> ‘to beat’, <i>contrastar</i> ‘to contrast’, <i>derrocar</i> ‘to overthrow’, <i>rechinar</i> ‘to grind’, <i>tronar</i> ‘to thunder’, <i>vocear</i> ‘to holler’ | 1 | –2334 |
| 19th | <i>disparatar</i> ‘to talk nonsense’, <i>mecer</i> ‘to rock’, <i>refregar</i> ‘to scrub’, <i>regañar</i> ‘to scold’, <i>tragar</i> ‘to swallow’, <i>zaherir</i> ‘to taunt’ | 1 | –1980 |
| 20th | <i>acumular</i> ‘to accumulate’, <i>cegar</i> ‘to blind’, <i>chancearse</i> ‘to jest’, <i>comerciar</i> ‘to trade’, <i>dejar</i> ‘to leave’, <i>insultar</i> ‘to insult’, <i>maldecir</i> ‘to curse’, <i>parlotear</i> ‘to chatter’, <i>penar</i> ‘to suffer’, <i>preparar</i> ‘to prepare’, <i>rebullir</i> ‘to start moving’, <i>retroceder</i> ‘to move back’, <i>roznar</i> ‘to bray’, <i>soltar</i> ‘to release’, <i>trajinar</i> ‘to bustle’, <i>trapear</i> ‘to mop’ | 1 | –1531 |
| 20th | <i>nevar</i> ‘to snow’, <i>provocar</i> ‘to provoke’, <i>quejarse</i> ‘to complaint’, <i>rezongar</i> ‘to grumble’ | 2 | –3065 |

Table 10 shows the lexemes that, statistically, prefer *parar de* in a certain period, although we only documented one occurrence of them in the construction. The frequency column indicates the number of occurrences of *dejar de* in that period, for comparison.

Table 10. Lexemes that only occur with *parar de* once in certain periods but prefer it.

| Period | Lexeme | Frequency with <i>dejar de</i> | collStr |
|------------|--|--------------------------------|---------|
| until–16th | <i>acrecentar</i> ‘to increase’ | 2 | –2423 |
| | <i>revolver</i> ‘to stir’ | 3 | –2298 |
| | <i>tañer</i> ‘to toll’ | 14 | –1727 |
| 17th–18th | <i>danzar</i> ‘to dance’, <i>recoger</i> ‘to pick up’ | 1 | –2034 |
| | <i>moler</i> ‘to grind’, <i>porfiar</i> ‘to persist’ | 3 | –1735 |
| | <i>agradecer</i> ‘to thank’, <i>combatar</i> ‘to fight’ | 6 | –1495 |
| | <i>tentar</i> ‘to tempt’, <i>volar</i> ‘to fly’ | 7 | –1438 |
| 19th | <i>bullir</i> ‘to boil’, <i>exclamar</i> ‘to exclaim’, <i>hervir</i> ‘to boil’, <i>subir</i> ‘to go up’, <i>vomitara</i> ‘to throw up’ | 1 | –1681 |
| | <i>añadir</i> ‘to add’ | 2 | –1507 |
| | <i>practicar</i> ‘to practice’ | 3 | –1384 |

Finally, Table 11 shows the lexemes that, statistically, prefer *parar de* in a certain period, occur in that construction more than once, and also occur with *dejar de*. The frequency column indicates the occurrences with *parar de* vs. those with *dejar de*.

Figure 3 below presents the level of attraction of lexemes with significant attraction to any of the two constructions over time. Each line represents a lexeme that prefers either *parar de* or *dejar de* in any period. In order to show the evolution of the lines more clearly, we show each century, from the 17th century onwards, separately. If a lexeme is positioned between the dotted lines, the preference is not significant. The most striking lexemes, with extraordinary preference for either construction, are indicated with the corresponding label: *ser* ‘to be’ in the case of *dejar de*, and *contar* ‘to tell’ in the case of *parar de*. Interestingly enough, the combination of *parar* with the predicate *contar* is so frequent that it has ended up being lexicalized in the locution *y para/pare de contar* (lit. ‘and stop of to tell/count’) used to put an end to an enumeration or narration.

Table 11. Lexemes that prefer *parar de* in certain periods and occur more than once.

| Period | Lexeme | Frequency with <i>parar de</i> vs. <i>dejar de</i> | collStr |
|------------|-------------------------------|--|---------|
| until-16th | <i>trabajar</i> 'to work' | 2 vs. 52 | -2690 |
| 17th-18th | <i>hablar</i> 'to speak' | 3 vs. 44 | -2890 |
| | <i>beber</i> 'to drink' | 2 vs. 11 | -2802 |
| 19th | <i>contar</i> 'to tell' | 19 vs. 13 | -30,602 |
| | <i>correr</i> 'to run' | 5 vs. 17 | -5617 |
| | <i>echar</i> 'to throw' | 4 vs. 19 | -4084 |
| | <i>clamar</i> 'cry out' | 2 vs. 1 | -3493 |
| | <i>hablar</i> 'to speak' | 4 vs. 60 | -2366 |
| 20th | <i>contar</i> 'to tell' | 35 vs. 12 | -44,326 |
| | <i>remover</i> 'to stir' | 10 vs. 3 | -12,986 |
| | <i>hablar</i> 'to speak' | 18 vs. 85 | -9108 |
| | <i>llover</i> 'to rain' | 10 vs. 23 | -7700 |
| | <i>hervir</i> 'to boil' | 5 vs. 5 | -5329 |
| | <i>revolver</i> 'to stir' | 4 vs. 5 | -4088 |
| | <i>dar</i> 'to give' | 11 vs. 94 | -3618 |
| | <i>cocer</i> 'to cook' | 3 vs. 4 | -3094 |
| | <i>beber</i> 'to drink' | 5 vs. 31 | -2424 |
| | <i>protestar</i> 'to protest' | 2 vs. 2 | -2304 |
| | <i>gritar</i> 'to shout' | 3 vs. 10 | -2239 |
| | <i>decir</i> 'to say' | 5 vs. 37 | -2133 |
| | <i>subir</i> 'to go up' | 2 vs. 3 | -2090 |
| | <i>tocar</i> 'to touch' | 5 vs. 55 | -1509 |
| | <i>bailar</i> 'to dance' | 3 vs. 22 | -1446 |

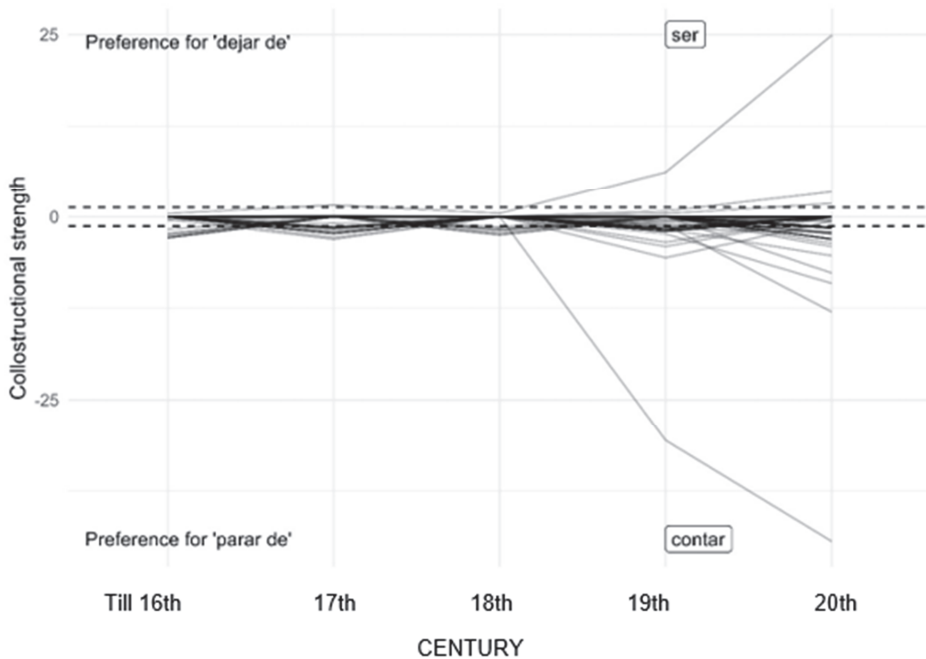


Figure 3. Level of attraction of lexemes with significant attraction to any of the two constructions over time.

As these two extraordinary lexemes dominate the graph in Figures 3–5, focus on smaller values, highlighting and labeling lexemes that prefer either one construction or the other in the first and last periods, i.e., until–16th century and the 20th century, respectively, so as to portray the evolution of the preferences.

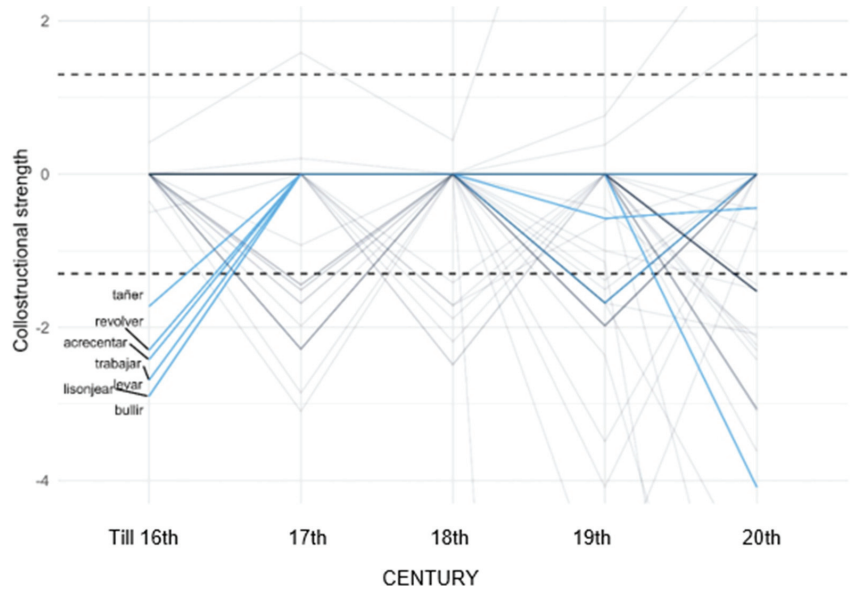


Figure 4. Level of attraction of lexemes with significant attraction to a construction up to the 16th century, throughout time.

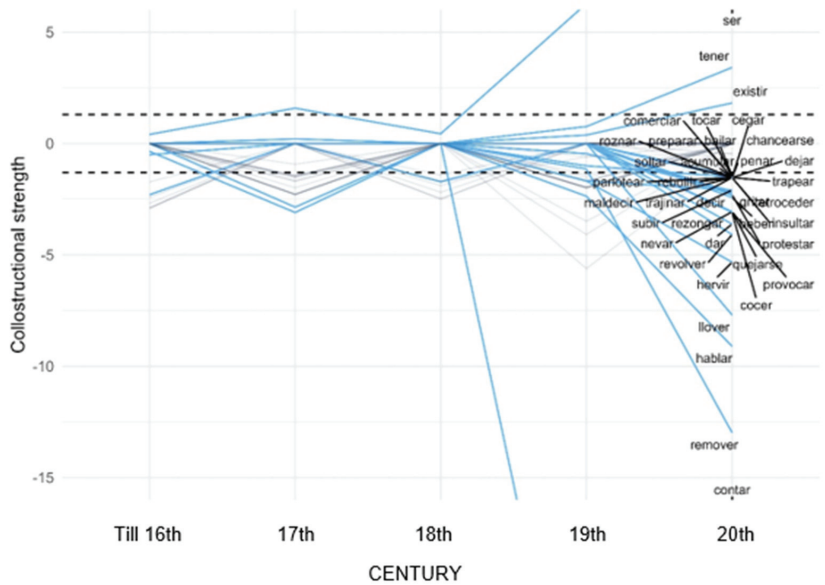


Figure 5. Level of attraction of lexemes with significant attraction to any of the constructions in the 20th century, throughout time.

Figure 4 is focused on the 16th century, a key period for the history of the two periphrases under study since it is the moment in which *parar de + INF* is defined as a construction in Spanish. At this time, as is to be expected given its recent introduction into the language, this periphrasis shows a particular affinity for certain verbal forms that express processes or achievements, such as *tañer* ‘to toll,’ *revolver* ‘to stir,’ *acrecentar* ‘to increase,’ *trabajar* ‘to work,’ *llevar* ‘to carry,’ *lisonjear* ‘to flatter’ and *bullir* ‘to boil.’ By contrast, the most well-established construction in the language, *dejar de + INF*, shows no particular constructional preference.

Figure 5 is focused on the use of *dejar de + INF* and *parar de + INF* in the 20th century. Unlike what we saw in Figure 4, Figure 5 shows a marked tendency for the first of these periphrases to be used with states (specifically, with three verbs: *ser* ‘to be,’ *tener* ‘to have,’ *existir* ‘to exist’). In turn, *parar de + INF* shows greater constructional diversity, which is reflected in the dense network of combinations made up by the different verbs that appear in the Vinf slot and that show a statistically significant attraction to the construction. Notice that the label that goes with the upper line corresponds to *ser* ‘to be,’ which is strongly attracted to *dejar de + INF*, and the one that goes with the lower line corresponds to *contar* ‘to tell,’ which is strongly attracted to *parar de + INF*. These are the two extraordinary lexemes that dominate the graph in Figure 3 above.

Once again, the collocation study per period offers data that coincides with those resulting from the analysis of the *Aktionsart* of the verbs that appear in the Vinf slot of the periphrases *dejar de + INF* and *parar de + INF*. In addition, these results allow for drawing relevant conclusions on the semantic evolution of both periphrases. As can be seen in Figure 6, if we pay attention to the distribution of verb types by centuries, we observe that processes and accomplishments dominate almost exclusively throughout all evolutionary stages of *parar de + INF*.

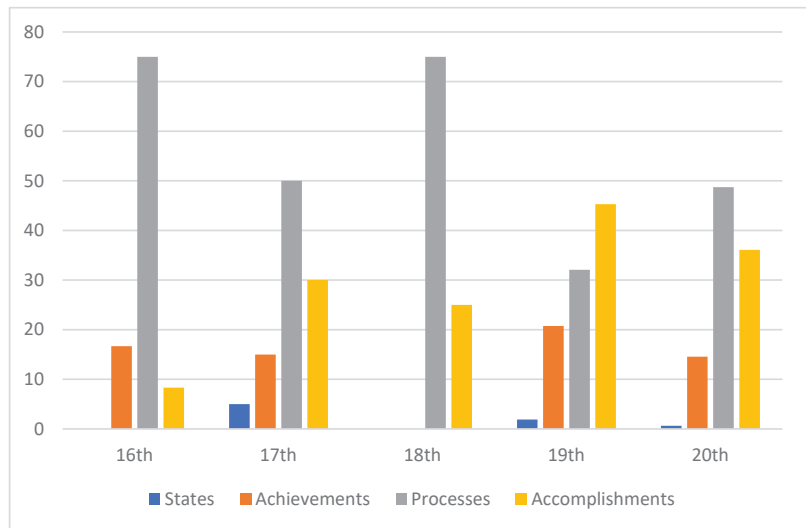


Figure 6. *Aktionsart* of the infinitive in the periphrasis *parar de + INF* throughout time.

Unlike *parar de + INF*, *dejar de + INF* shows a tendency to combine, always exclusively, with processes and accomplishments up to the 15th century, where we observe a clear rise of achievements, as Figure 7 shows. From the 16th century onward, achievements and states exceeded processes and accomplishments, with a growing disproportion from the 17th century onward. This situation continued until contemporary times, despite the fact that, in the 20th century, a recovery of the processes was documented.

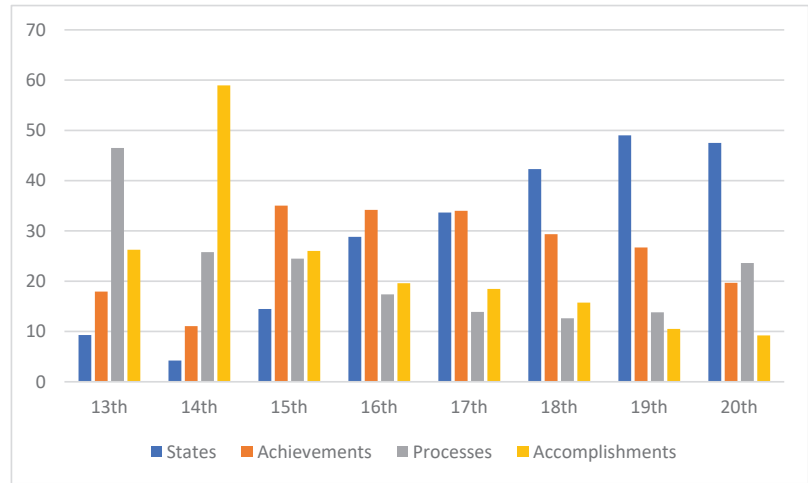


Figure 7. Aktionsart of the infinitive in the periphrasis *dejar de + INF* throughout time.

The collocation analysis, together with the study of the Aktionsart of the verbs that appear in the infinitive position, allowed us to draw interesting conclusions about the emergence of *parar de + INF* in Spanish as well as about the possible analogical influence that *dejar de + INF* could have exerted. The interruptive meaning of *dejar de + INF*, the most common in the medieval period, could have attracted the construction with *parar*, which might also have experienced the influence of other interruptive and continuative periphrases such as *cesar de + INF*. However, the results offered by the collocation analysis and the analysis of the Aktionsart of the predicates in the Vinf slot show that this influence had to occur at a schematic level since the combinatorial patterns of *parar de + INF* present it as a construction with its own identity, which does not replicate the model of the most frequent interruptive periphrasis in Spanish. If the influence had occurred at the most saturated level of the construction, *parar de + INF* would have replicated the collocation patterns of *dejar de + INF*, and no significant differences would be detected between the two periphrases regarding the type of verb that appears in the non-finite position.

On the collocation side, we observe that already in the 16th century, at the moment when *parar de + INF* emerges in Spanish, there are lexemes that are only documented with this construction (Table 9) or that prefer it (Tables 10 and 11). Therefore, we can claim that the emergence of *parar de + INF* in Spanish does not follow the collocation patterns of *dejar de + INF*. On top of that, in the 16th century, this construction did not follow the model of *dejar de + INF* in terms of the Aktionsart of the verbs occupying the Vinf slot. As shown in Figure 8, even at this time in the history of Spanish, both periphrases show clearly differentiated patterns.

The preference of *parar de + INF* for processes and accomplishments, attested from its earliest documentation, makes it possible to claim that *parar de + INF* only competes with *dejar de + INF* in their uses as phasal periphrases, given that the interruptive and continuative meanings have always characterized the periphrasis; cf. Figure 9, adapted from Garachana (2021).

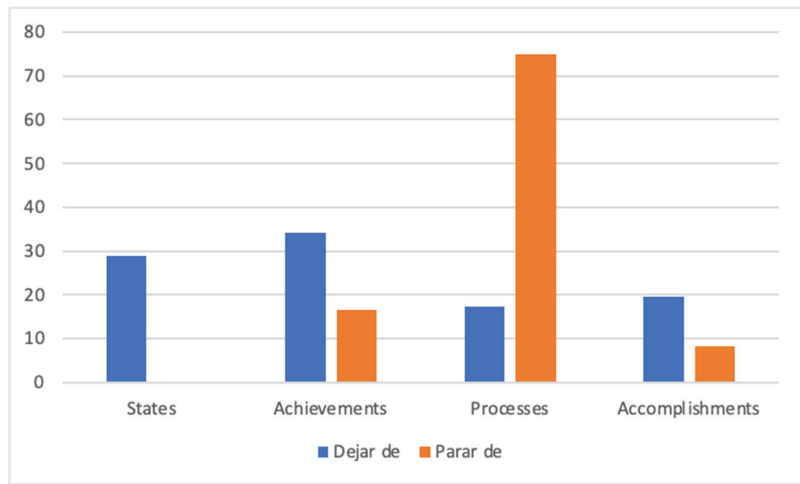


Figure 8. Aktionsart of the Vinf of both periphrases in the 16th century.

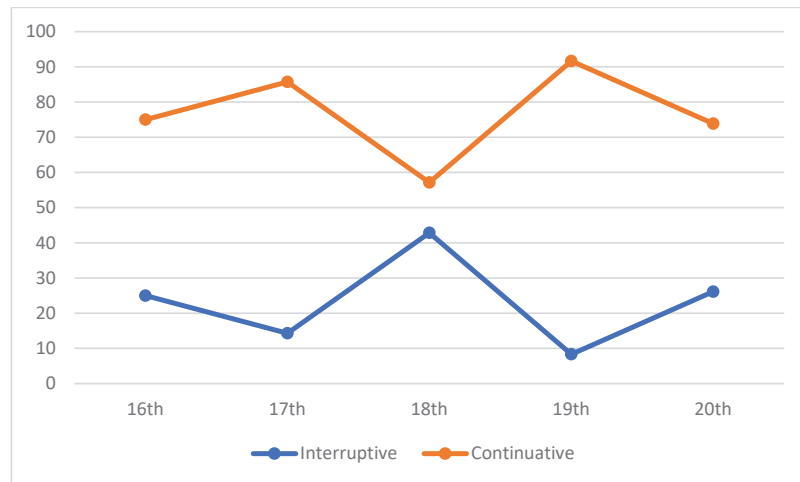


Figure 9. Values of *parar de* + INF throughout the history of Spanish (16th–20th centuries).

Interestingly, after the introduction of *parar de* + INF in the paradigm of periphrases, a significant change in the use of *dejar de* + INF occurs, as shown in Figure 10. In fact, starting in the 16th century, we observe the collapse of the interruptive meanings that, during the medieval period, had been the most characteristic of this periphrasis. From the 16th century on, the periphrasis was mainly used with assertive values or to deny the event expressed by the infinitive. This semantic evolution is linked to the regression of atelic predicates in the Vinf slot and allows us to advance the hypothesis that, since there is a new piece in the system of interruptive periphrases, which was added to the already consolidated *cesar de* + INF (see Yllera 1980 for more information), this would allow *dejar de* + INF to expand towards other areas of meaning, linked to stylistic and discursive needs. In the same way, the recovery of the interruptive meanings of the periphrasis in the 20th century is related to the increase in processes and can be explained by an analogical influence on the part of *parar de* + INF at the moment in which this periphrasis is consolidated in the language.

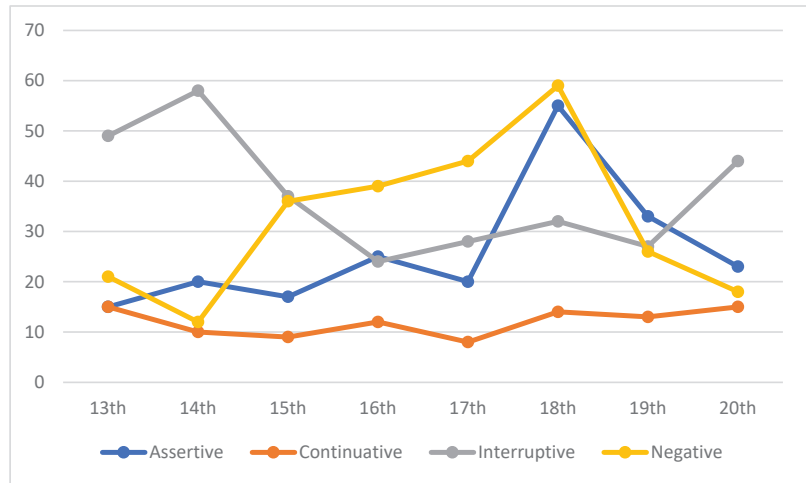


Figure 10. Values of *dejar de + INF* throughout the history of Spanish.

5. Discussion

The findings of the analysis of the productivity of both periphrastic constructions, as well as those that emerge from the collostructional analysis carried out, show that *dejar de + INF* and *parar de + INF* behave in particular ways that do not coincide with the descriptions contained in the existing literature on these periphrases. As was stated in the introduction, grammars readily admit the periphrastic character of *dejar de + INF*, while *parar de + INF* is usually subject to interpretations that oscillate between those who accept its status as a periphrasis and those who consider that it functions as a semi-auxiliary verb. However, the analysis of the use of these periphrases based on an extensive corpus forces us to review these statements.

The data show that the realized and potential productivity of *parar de + INF* is higher than that of *dejar de + INF*. In fact, the realized productivity of *dejar de + INF* is always lower than that of *parar de + INF*. And if we compare the contemporary realized productivity of *parar de + INF* (0.37) with that which *dejar de + INF* presented in the medieval period (0.23 in the 13th–15th centuries; 0.29 in the 14th century), that of *parar de + INF* is higher. Similarly, the potential productivity of *parar de + INF* in the 20th century (0.22) remains above that of *dejar de + INF* in the medieval period (0.12 in the 13th–14th centuries; 0.19 in the 14th century), when its use in the language must still have been recent and an expansion of the periphrasis could be expected.

As noted earlier (cf. Section 4.2), we could argue that this behavior is due to a more advanced stage of grammaticalization of *dejar de + INF*, which already in the medieval period would have advanced more in the language than *parar de + INF*. Indeed, the most grammaticalized periphrases have reached higher levels of lexical expansion than those that are not yet established in the language; an incipient grammaticalization determines that the possibilities of expansion turn out to be greater. Once a periphrasis is fixed for the expression of temporal, aspectual, or modal values, it usually presents few restrictions regarding the verb types that can appear in the infinitive position. If they exist, such restrictions have to do with the meaning that the periphrasis expresses. However, in the case of *dejar de + INF*, perhaps this lower productivity may be due to other grammatical and discursive motivations, especially if we consider that, according to Garachana (n.d.), this periphrasis shows few signs of grammaticalization throughout the history of Spanish, and its use seems to be closely linked to particular stylistic and rhetorical motivations from early dates.

Interestingly, Yllera (1980) already highlights it as the most frequent of the medieval interruptive periphrases (Yllera 1980, p. 194) and points out its scarce employment with continuative value (Yllera 1980, p. 206). This behavior separates *dejar de* + INF from the rest of the medieval interruptive periphrases mentioned in Yllera (1980), since these were mostly used in negative modality sentences to denote the continuity of the action expressed by the infinitive. In addition, unlike those other periphrases, *dejar de* + INF could be used to indicate the negation of the event expressed by the infinitive (Yllera 1980, p. 215) and to formulate more or less forceful assertions.

In this way, *dejar de* + INF was already outlined in the Middle Ages as a periphrasis with its own defining characteristics, which do not fit what is indicated in the bibliography on the subject. Along the same lines, its preference for combining with infinitives that express either achievements or states is contrary to what is expected of an interruptive (or continuative) periphrasis. Certainly, often the use of verbs that express achievements or states in the infinitive position of *dejar de* + INF allows adding a notion of repetition or iteration of the event (see Section 3); however, in most cases the construction expresses an assertion or negates the event expressed by the infinitive. To this propensity to combine with verbs that do not express actions that can be interrupted or continued, we must add the fact that our collocation analysis clearly indicates that the lexemes that prefer *dejar de* + INF are above all states (specifically, these are the verb forms *ser* ‘to be,’ *tener* ‘to have,’ *haber*, ‘to have’). Among the accomplishments and the processes, only *hacer* ‘to do’ and *ver* ‘to see’ show a tendency to be used in the periphrasis.

It is interesting that the link between *dejar de* + INF and concrete verb types began in the 17th century—remember that, prior to that century, no verb form seemed to show a particular preference for this periphrasis—, since, in principle, a grammaticalized periphrasis should have a somewhat homogeneous behavior with all kinds of verb forms. However, starting in the 17th century, *dejar de* + INF began to specialize on a few verb types, which gave it a certain formulaic quality, atypical of a verbal periphrasis. On the other hand, *parar de* + INF shows that the verbs that prefer it are much more varied (vid. Table 6), and this group of verbs even includes meteorological verbs, which is usually considered a proof of the grammaticalization of verbal periphrases (Garachana 2017, p. 43). Furthermore, if we observe the collocation preferences of *parar de* + INF per period, we observe that the number of verb forms increases throughout time. That is, *parar de* + INF shows greater collocation flexibility than *dejar de* + INF.

Thus, the collocation behavior of *parar de* + INF and *dejar de* + INF offers contrasting results regarding the very consideration of these constructions as verbal periphrases. The results of our analysis show that the form that best fits the definition of verbal periphrasis is, paradoxically, the construction that has given rise to the greatest debate about its periphrastic or non-periphrastic character and status. Indeed, *parar de* + INF has higher realized and potential productivity than *dejar de* + INF and manifests fewer restrictions that affect the types that can combine with it. On the other hand, *dejar de* + INF appears to be a more rigid construction than *parar de* + INF, with a certain tendency towards routinization with particular verb types.

At this point, it is worth asking why *dejar de* + INF tended towards routinization starting precisely in the 17th century. The answer probably lies in rhetorical and stylistic preferences. Precisely, this was the moment in which there was a higher use of periphrases within auxiliary chains, for instance, ‘*No puedo dejar de admitir*’, ‘I must (lit. cannot stop) admitting’. As Garachana (n.d.) shows, in this century the use of periphrastic chains in CORDE amounts to 18.50% of the total number of uses of verbal periphrases, and the percentages for subsequent centuries are: 13.18% in the 18th century, 12.05% in the 19th century, and 9.55% in the 20th century. Significantly, within these auxiliary chains, the most frequent verb is *ser*, ‘to be,’ which, in the 17th century, appeared in 13.66% of the periphrastic chains. This percentage rises to 21.27% and 21.83%, respectively, in the 18th and 19th centuries, and then drops to 13.33% in the 20th century (Garachana n.d.). These findings show that in a significant number of cases, the use of the periphrasis is associated

with a more complex construction in which the verb *ser*, ‘to be’ occupies a prominent place. Now, this chained construction does not express aspectual meanings but negates the meaning of the event expressed by the infinitive.

These values are usually not the most frequent ones for verbal periphrases in Spanish, which are usually associated with the expression of temporal, aspectual, or modal meanings. Moreover, *dejar de* + INF is frequently used for the expression of assertive, commissive, and directive speech acts. At this point, it is worth asking whether *dejar de* + INF can really be considered, in all its meanings, a representative member of the category of verbal periphrases. Its restrictions related to its productivity, the strong affinity it shows for only a few verb forms, and its propensity to be used in auxiliary chains place it on the margins of the category of periphrasis. Only its uses as an interruptive and continuative periphrasis allow its inclusion in the category as a full member. The rest of its values force us to reflect on the type of category to which it should be assigned.

This issue goes beyond the discussion about the grammaticalization of verbal constructions since it affects the values they express. The most recent debates on this matter (see García Fernández et al. 2006; Carrasco Gutiérrez 2008; García Fernández and Carrasco Gutiérrez 2008) propose characterizing the constructions *acabar* + GER, *acabar por* + INF, *alcanzar a* + INF, *comenzar* + GER, *comenzar por* + INF, *empezar* + GER, *empezar por* + INF, *llegar a* + INF, *pasar a* + INF, *terminar* + GER, *terminar por* + INF, and *venir a* + INF as discourse markers. The decisive argument of these authors, counterargued by Olbertz (2007), rests on the fact that these periphrases do not express temporal, aspectual, or modal meanings, as is characteristic of the category, but rather, on the contrary, function as discourse structuring devices. Contrarily, Olbertz (2007) emphasizes the formal features of these periphrases to defend their maintenance within the category. In our opinion, both Carrasco and García Fernández, as well as Olbertz, are correct. On the one hand, the aforementioned periphrases work with values that are more typical of information structuring devices than those of verbal periphrases. On the other hand, the grammatical functioning of these structures is characteristic of verbal periphrases. Thus, if we also consider the characterization of *dejar de* + INF advanced here, it is necessary to review the very concept of periphrasis and to qualify the types of meanings that can be expressed by these grammatical constructions.

Considering all of the above, the evolution of these two periphrases in Spanish shows a completed process of constructionalization since their first documentation. The process, however, seems to have advanced further in the case of *dejar de* + INF, which has reached a stage in which it has surpassed the expression of aspectual meanings to encompass pragmatic meanings linked to the expression of assertive, commissive, and directive speech acts. The evolution of *parar de* + INF would have been stopped in the phasal (interruptive or continuative) meanings.

As Table 12 shows, the constructionalization of *dejar de* + INF starts from the lexical transitive verb *dejar*, ‘to leave.’ This verb subcategorizes a direct object and a locative complement and means the action of depositing an object at a certain location, as in *Dejar algo en un sitio*, ‘place something somewhere’ (STAGE 1). Although there are documentations of the use of *dejar* as an auxiliary verb already in the 13th century, the frequent use of the locution *Dexa aquí la estoria de fablar en X*, ‘the narration of a certain matter is interrupted’ leads Garachana (n.d.) to propose the hypothesis that the periphrasis originates from such a locution, in which the verb *dejar* has a metaphorical meaning. Indeed, in *Dexa aquí la estoria de fablar en ello, dexa*, ‘present indicative of the verb *dejar*’ means ‘the abandonment of a story at a certain point in the narrative.’ It is, then, a medieval structure for changing the topic. This locution would have functioned as a bridging context: in it, *dejar* no longer requires an animate subject, and it often subcategorizes a deictic adverb (*aquí* ‘here’) that indicates the place in the story where it is abandoned, and what is deposited (or abandoned) metaphorically in that place is no longer an object (STAGE 2). From these bridging contexts, in which the interruption of a story is expressed, the interruptive aspectual meaning of the construction would have been reached (STAGE 3), where *dejar* functions as an auxiliary

verb, combined mainly with processes and accomplishments in the Vinf slot, and the structure *dejar de* + INF is already fully constructionalized.

Table 12. The constructionalization process of *dejar de* + INF.

| | Lexicon | | Grammar | | Pragmatics |
|--------------------------|--|--|---|--|--|
| <i>Dejar de</i> + INF | STAGE 1 Etymon | STAGE 2 Bridging context | STAGE 3 Interruptive Verbal Periphrasis | STAGE 4 Continuative Verbal periphrasis | STAGE 5 Assertive and negative values |
| Semantics/ pragmatics | <i>dejar</i> = ‘place sth. somewhere’ <i>Dejó los libros sobre la mesa</i> | Old Spanish: ‘interrupt (stop) a story at a point’ <i>Dexa aquí la estoria de fablar</i> | ‘to interrupt an event’ <i>Dejó de sufrir</i> | ‘to maintain an event’ <i>No dejó de fumar ni un momento</i> | ‘To say or do something’ <i>No debes de ir</i> <i>No puedo dejar de admitir</i> |
| Syntax | subject [+animate] + <i>dejar</i> + DO + locative complement | Subject [+/- animate] + <i>dejar</i> + locative complement (\emptyset /a/de) + PP (de/a INF) | <i>dejar de</i> + INF | NEG <i>dejar de</i> + INF | AFF/NEG <i>dejar de</i> + INF Auxiliary verb + <i>dejar de</i> + INF |
| | | S XIII & XIV. INF that mainly express ‘telling a story’ (<i>decir</i> ‘to say’, <i>contar</i> ‘to tell’, <i>hablar</i> ‘to speak’) | Other types start consolidating (processes, accomplishments, non-permanent states) | Processes, accomplishments, non-permanent states | INF = achievement or state Often, a temporal expression delimits the duration of the infinitive. <i>Dejar de</i> + INF is often preceded by another auxiliary verb (auxiliary chains) |

However, the evolution did not stop at this point, since when using *dejar de* + INF in contexts of negative modality, it acquires a continuative meaning (STAGE 4). Later, the periphrasis will be used to express assertions and the negation of the event expressed by the infinitive, usually in the formulation of assertive, commissive, or directive speech acts. The availability of these values presupposes a formal change in the construction, as they are triggered when the Vinf slot is occupied by states and achievements, often in auxiliary chains. In addition, often the event expressed by the verb is temporally delimited by a temporal expression (STAGE 5). Table 12 summarizes this constructionalization process, which starts in the lexicon and concludes in pragmatics.

The constructionalization of *parar de* + INF is represented in Table 13 below. The starting point is found in the verb *parar*, a form that in the medieval period expressed both locative values (*parar* meant ‘to place something in a position’) and intransitive senses (*parar* = to stop) (Corominas and Pascual 1991: s.v. *parar*). In its locative senses, *parar* was close to *dejar*, which could have contributed to the use of *parar* as an auxiliary verb. It has been proven that in Spanish, it is common for the lexical synonymy between two verbs—one of which also functions as an auxiliary verb—to originate processes of analogical attraction that conclude with the auxiliary use of the other verb (Garachana 2011; Garachana and Rosemeyer 2011). This process would have been activated by the intransitive sense of *parar*, which expresses the interruption of a movement and, therefore, is linked to the interruptive aspectual sense that the periphrasis expresses in certain contexts. As in the case of *dejar de* + INF, the use of *parar de* + INF in negative modality contexts would motivate the development of continuative meanings.

Table 13. The constructionalization process of *parar de* + INF.

| | Lexicon | | Grammar | |
|-----------------------|---|---------------------------------------|--|--|
| <i>Parar de</i> + INF | STAGE 1 Etymon | STAGE 2 Supporting construction | STAGE 3 Interruptive Verbal periphrasis | STAGE 3–4 Continuative Verbal periphrasis |
| Semantics/pragmatics | <i>parar</i> ‘put in a position’ > ‘to stop physically’ <i>Los comeres delant gelos</i> <i>paravan</i> (Cid) <i>Paravas delant al campeador</i> | <i>Dejar de</i> + INF | ‘to interrupt an event’ (nuance of annoyance; directive speech acts) <i>Para de moverte</i> | ‘to maintain an event’ (nuance of annoyance) <i>No paran de malestar</i> |
| Syntax | <i>parar</i> (tr. V.) + OD + locative complement <i>parar</i> (intr. V.) + locative complement | | <i>parar de</i> + INF | NEG <i>parar de</i> + INF |

6. Conclusions

The main contribution of the preceding discussion lies in having verified our hypothesis that the introduction of *parar de* + INF in the paradigm of verbal periphrases did not imply a copy of existing models but rather that the periphrasis followed its own constructional pattern. In addition, we were able to confirm that this introduction of *parar de* + INF into the system of verbal periphrases had collateral effects on the use of *dejar de* + INF, which, in turn, came to be used with values less linked to phasal meanings, which had been its defining values during the medieval period.

The conclusions that can be drawn for the study of *parar de* + INF and *dejar de* + INF are highly relevant and connect with findings of previous productivity research that has focused on quantitative frequency measures (see Hilpert 2013; Perek 2020). The study of their normalized frequency of use and of their realized and potential productivity, as well as the collostructional analysis carried out, allow us to claim that *parar de* + INF is consolidating in the language as a productive periphrasis, which attracts an ever-growing group of verb forms in the non-finite verb slot. In addition, its meaning is associated with the aspectual values that are typical of interruptive and continuative periphrases. On the other hand, *dejar de* + INF from the medieval period onwards appears as an increasingly less productive and more rigid construction than *parar de* + INF, with a certain tendency towards routinization with particular verb types. In addition, it is gradually being less and less used with interruptive aspectual values (continuatives were never especially frequent), and it is emerging as a construction specialized in the expression of values that are not typical of verbal periphrases. As in the case of the periphrases analyzed by García Fernández and Carrasco Gutiérrez (2008), we find, under a periphrastic form, a construction capable of expressing pragmatic meanings associated with a specific stylistic and communicative will. Thus, despite the fact that *parar de* + INF is the form that has aroused the greatest reluctance in the literature on periphrases when it comes to including it as a member of the category, an in-depth study based on a large corpus reveals that this construction fits the characteristics of the category even better than *dejar de* + INF does.

Still, some open questions remain. We wonder about the impact that the type of corpus used has had when it comes to describing the evolution and frequency of the periphrases. The use of *parar de* is associated with the expression of directive speech acts as well as of annoyance, as in *¡Para de moverte ya!* ‘Stop moving already!’. We could assume that the low token frequency of *parar de* we observed may be partially explained by the kind of corpus that we used, which was only composed of written documents, which tend to avoid situations of communicative proximity. This is obviously only relevant for the centuries for which we have oral data available; however, it would be interesting to look into more recent oral corpora to check whether the observed distribution holds. This disadvantage can be remedied by complementing methodological approaches by, for instance, carrying out experiments such as acceptability rating or sentence completion tasks in order to investigate the extensibility of a construction in more detail. In future studies, we would also like to

focus more on the sematic productivity of *parar de* + INF and *dejar de* + INF by conducting a fine-grained analysis of the semantic characteristics of the non-finite verbs in these verbal periphrases following the ADESSE classification (see García Miguel et al. 2010). Finally, we are planning on looking more closely at the kind of attraction that near-synonymous constructions exert to explain how the Spanish periphrastic system became structured to prevent constructions from developing the exact same meaning and how it is being recomposed throughout time.

Author Contributions: Conceptualization, M.G. and M.S.S.; methodology, M.G. and M.S.S.; software, M.G. and M.S.S.; validation, M.G. and M.S.S.; formal analysis, M.G. and M.S.S.; investigation, M.G. and M.S.S.; resources, M.G. and M.S.S.; data curation, M.G. and M.S.S.; writing—original draft preparation, M.G. and M.S.S.; writing—review and editing, M.G. and M.S.S.; visualization, M.G. and M.S.S.; supervision, NA; project administration, M.G. and M.S.S.; funding acquisition, M.G. and M.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project *Diccionario histórico de las perifrasis verbales del español. Gramática, pragmática y discurso (III). Perifrasis formadas a partir de verbos superléticos y de movimiento* of MINECO (PID2019-105415RB-I00) and by the C14/20/038 project *Speech-act orientatie en intersubjectief alignment*, financed by the Research Council of the KU Leuven.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The corpus data presented in this study are available online (CORDE). The analyzed data are available upon request from the corresponding authors.

Acknowledgments: The authors would like to acknowledge the technical support given by Mariana Montes (KU Leuven) with regards to the distinctive collexeme analysis. Likewise, the authors would like to thank Daniel Cuní Díez and Samia Aderdouch for their help in the process of data compilation.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Aaron, Jessi E. 2010. Pushing the envelope: Looking beyond the variable context. *Language Variation and Change* 22: 1–36. [CrossRef]
- Aparicio Mera, Juan J. 2016. Representación computacional de las perífrasis de fase: De la cognición a la computación. Ph.D. dissertation, University of Barcelona, Barcelona, Spain. Available online: https://diposit.ub.edu/dspace/bitstream/2445/101660/1/Tesis_JJAparicio.pdf (accessed on 15 December 2021).
- Baayen, Harald R. 2009. Corpus linguistics in morphology: Morphological productivity. In *Corpus Linguistics. An International Handbook*. Edited by Anke Lüdeling and Merja Kytö. Berlin: Mouton De Gruyter, pp. 900–19. [CrossRef]
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.
- Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer, and Spike Gildea. 2015. *Diachronic Construction Grammar*. Amsterdam: John Benjamins.
- Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In *The Handbook of Historical Linguistics*. Edited by Brian Joseph and Richard Janda. Oxford: Blackwell, pp. 602–23.
- Bybee, Joan, and Rena Torres-Cacoullous. 2009. The role of prefabs in grammaticization. How the particular and the general interact in language change. In *Formulaic Language. Volume 1. Distribution and Historical Change*. Edited by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley. Amsterdam and Philadelphia: John Benjamins, pp. 187–217.
- Camus, Bruno. 2006a. Dejar de + infinitivo. In *Diccionario de perifrasis verbales*. Edited by García Fernández Luis, Ángeles Carrasco Gutiérrez, Bruno Camus, María Martínez Atienza and María Ángeles García. Madrid: Gredos, pp. 117–20.
- Camus, Bruno. 2006b. Parar de + infinitivo. In *Diccionario de perifrasis verbales*. Edited by García Fernández Luis, Ángeles Carrasco Gutiérrez, Bruno Camus, María Martínez Atienza and María Ángeles García. Madrid: Gredos, pp. 206–9.
- Carrasco Gutiérrez, Ángeles. 2008. *Llegar a + infinitivo* como conector aditivo en español. *Revista española de lingüística* 38: 67–94. [CrossRef]
- Copple, Mary T. 2011. Tracking the constraints on a grammaticalizing perfect(ive). *Language Variation and Change* 23: 163–91. [CrossRef]
- Corominas, Joan, and José Antonio Pascual. 1991. *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.
- Croft, William. 2000. *Explaining Language Change. An Evolutionary Approach*. Harlow: Longman.
- Croft, William, and Alan Cruise. 2004. *Cognitive Linguistics*. New York: CUP.
- Fernández de Castro, Félix. 1999. *Las perifrasis verbales en el español actual*. Madrid: Gredos.
- Fillmore, Charles J. 1996. The Pragmatics of Constructions. In *Social Interaction, Context, and Language*. Edited by Dan Isaac Slobin, Julie Gerhardt, Amy Kyratzis and Jiansheng Guo. New Jersey: L. Erlbaum Associates, pp. 53–69.

- Fried, Miriam. 2009. Construction Grammar as a tool for diachronic analysis. *Constructions and Frames* 1: 261–90. [CrossRef]
- Garachana, Mar. 2011. Perífrasis sinónimas. ¿Gramaticalizaciones idénticas? Mas retos para la teoría de la gramaticalización. In *Sintaxis y análisis del discurso hablado en español: Homenaje a Antonio Narbona*. Edited by José Jesús de Bustos Tovar, Rafael Cano Aguilar, Elena Méndez García de Paredes and Araceli López Serena. Sevilla: Universidad de Sevilla, pp. 779–98.
- Garachana, Mar. 2016. Restricciones léxicas en la gramaticalización de las perífrasis verbales. *Rilce* 32: 136–58. [CrossRef]
- Garachana, Mar. 2017. Los límites de una categoría híbrida. Las perífrasis verbales. In *La gramática en la diacronía. La evolución de las perífrasis verbales modales en español*. Edited by Mar Garachana Camarero. Madrid-Frankfurt: Iberoamericana-Vervuert, pp. 35–80. [CrossRef]
- Garachana, Mar. 2021. Bailando, me paso el día bailando y los vecinos mientras tanto no paran de molestar. *Parar de + Inf* as an Interruptive Verbal Periphrasis in Spanish. *Languages* 6: 171. [CrossRef]
- Garachana, Mar. 2022. Unexpected grammaticalizations. The reanalysis of the Spanish verb *ir* ‘to go’ as a past marker. In *From Verbal Periphrases to Complex Predicates*. Edited by Mar Garachana, Sandra Montserrat and Claus Pusch. Amsterdam: John Benjamins, pp. 171–88. [CrossRef]
- Garachana, Mar. n.d. Gramaticalización, tradicionalidad discursiva y presiones paradigmáticas en la evolución de dejar de + INF en español. *Beihefte zur Zeitschrift für romanische Philologie*. [CrossRef]
- Garachana, Mar, and Malte Rosemeyer. 2011. Rutinas léxicas en el cambio gramatical. El caso de las perífrasis deónticas e iterativas. *Revista de Historia de La Lengua Española* 6: 35–60. [CrossRef]
- García Fernández, Luis, and Ángeles Carrasco Gutiérrez. 2008. Perífrasis verbales con función de marcador del discurso. Contrarréplica a Olbertz (2007). *Verba* 35: 439–47.
- García Fernández, Luis, Ángeles Carrasco Gutiérrez, Bruno Camus, María Martínez Atienza, and María Ángeles García. 2006. *Diccionario de perífrasis verbales*. Madrid: Gredos.
- García Miguel, José M., Gael Vaamonde, and Fita González Domínguez. 2010. ADESSE, a Database with Syntactic and Semantic Annotation of a Corpus of Spanish. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*. Edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner and Daniel Tapias. Valetta: European Language Resources Association (ELRA), pp. 1903–10.
- Goldberg, Adele. 2006. *Constructions at Work. The Nature of Generalization in Language*. Oxford: OUP.
- Gómez Manzano, Pilar. 1992. *Perífrasis verbales con infinitivo: Valores y usos en la lengua hablada*. Madrid: UNED.
- Gómez Torrego, Leonardo. 1988. *Perífrasis verbales. Sintaxis, semántica y estilística*. Madrid: Arco Libros.
- Gómez Torrego, Leonardo. 1999. Los verbos auxiliares. Las perífrasis verbales de infinitivo. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa-Calpe, pp. 3323–89.
- Gries, Stefan, and Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics* 9: 97–129. [CrossRef]
- Hilpert, Martin. 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory* 2: 243–56. [CrossRef]
- Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge and New York: Cambridge University Press. ISBN 9781107013483.
- Hoffmann, Thomas. 2013. Abstract phrasal and clausal constructions. In *The Oxford Handbook of Construction Grammar*. Edited by Graeme Trousdale and Thomas Hoffmann. Cambridge: CUP, pp. 307–28. [CrossRef]
- Hoffmann, Thomas. 2014. The Cognitive Evolution of Englishes: The Role of Constructions in the Dynamic Model. In *The Evolution of Englishes: The Dynamic Model and Beyond*. Edited by Sarah Buschfeld, Thomas Hoffmann, Magnus Huber and Alexander Kautzsch. Amsterdam: John Benjamins, pp. 160–80.
- Hopper, Paul J. 1991. On some Principles of Grammaticalization. In *Approaches to Grammaticalization*. Edited by Elizabeth C. Traugott and Bernd Heine. Amsterdam: John Benjamins, vol. 1, pp. 17–35.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. 1. Theoretical Prerequisites*. Stanford: SUP.
- Langacker, Ronald W. 2003. Constructions in cognitive grammar. *English Linguistics* 20: 41–83. [CrossRef]
- Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Olbertz, Hella. 1998. *Verbal Periphrases in a Functional Grammar of Spanish*. Berlin: Walter de Gruyter.
- Olbertz, Hella. 2007. ¿Perífrasis verbales con función de marcador de discurso? A propósito del Diccionario de perífrasis verbales. *Verba* 34: 381–90.
- Perek, Florent. 2020. Productivity and schematicity in constructional change. In *Nodes and Networks in Diachronic Construction Grammar*. Edited by Lotte Sommerer and Elena Smirnova. Amsterdam and Philadelphia: John Benjamins, pp. 141–66.
- RAE/ASALE. 2009. *Nueva Gramática de la Lengua Española*. Madrid: Espasa.
- Real Academia Española. 2007. Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. Available online: <http://www.rae.es> (accessed on 12 July 2022).
- Rodríguez Molina, Javier. 2004. Difusión léxica, cambio semántico y gramaticalización: El caso de *haber* + participio en español antiguo. *Revista de filología española* 84: 169–209. [CrossRef]
- Rodríguez Molina, Javier. 2007. La selección del auxiliar en el Poema de Mio Cid y otros textos medievales: Cuestiones filológicas. In *Actes du XXIV Congrès International de Linguistique et de Philologie Romanes*. Edited by David. Trotter. Berlin: Max Niemeyer Verlag, vol. 2, pp. 299–310. [CrossRef]

- Rodríguez Molina, Javier, and Álvaro S. Octavio de Toledo y Huerta. 2017. La imprescindible distinción ente texto y testimonio: El CORDE y los criterios de fiabilidad lingüística. *Scriptum Digital* 6: 5–68.
- Rosemeyer, Malte, and Mar Garachana. 2019. De la consecución a la contraexpectación: La construccionalización de *Lograr/Conseguir + Infinitivo*. *Studies in Hispanic and Lusophone linguistics* 12: 383–422. [CrossRef]
- Torres Cacoullous, Rena. 2012. Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language* 36: 73–122. [CrossRef]
- Traugott, Elizabeth C., and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Yllera, Alicia. 1980. *Sintaxis histórica del verbo español: Las perifrasis medievales*. Zaragoza: Departamento de Filología Francesa, Universidad de Zaragoza.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

From Motion to Causation: The Diachrony of the Spanish Causative Constructions with *traer* ('Bring') and *llevar* ('Take')

Julio Torres Soler ^{1,*} and Renata Enghels ²

¹ Department of Spanish Studies, General Linguistics & Literature Theory, University of Alicante, 03690 Alicante, Spain

² Department of Linguistics, Ghent University, 9000 Ghent, Belgium

* Correspondence: julio.torres@ua.es

Abstract: This paper describes the historical evolution of the Spanish causative micro-constructions with the motion verbs *llevar* ('take') and *traer* ('bring') (e.g., *el miedo llevó al ladrón a cometer un error*, 'the fear caused the thief to make a mistake'). In order to reconstruct the historical development of these micro-constructions between the 13th and 20th centuries, all causative uses of *llevar* and *traer* were extracted from the *Corpus del Diccionario Histórico*. This corpus was annotated for a series of formal and semantic parameters that count as indexes of grammaticalization, and was submitted to a quantitative productivity analysis. The results point to the existence of a subschema formed of verbs of caused accompanied motion, which has semantically specialized in the expression of indirect causation. From a formal point of view, this subschema is characterized by a low level of syntactic incorporation of the causative verb and the infinitive. In addition, it is shown that the productivity of the causative micro-constructions under study is determined by semantic changes experienced by *llevar* and *traer* as full lexical verbs during the history of Spanish. The late development of the micro-construction with *llevar* is explained by the initial tendency of this verb to express motion events not bounded by an endpoint. From the 16th century onwards, the decline in the micro-construction with *traer* and the rise in the micro-construction with *llevar* results from the consolidation of the deictic meaning of the verb pair.

Keywords: grammaticalization; diachronic construction grammar; causation; causative construction; *llevar*; *traer*

Citation: Torres Soler, Julio, and Renata Enghels. 2023. From Motion to Causation: The Diachrony of the Spanish Causative Constructions with *traer* ('Bring') and *llevar* ('Take'). *Languages* 8: 122. <https://doi.org/10.3390/languages8020122>

Academic Editor: Jeanine Treffers-Daller

Received: 24 February 2023

Revised: 14 April 2023

Accepted: 17 April 2023

Published: 30 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

From a typological point of view, causation can be expressed by different linguistic procedures, including lexical, morphological and syntactic mechanisms (Shibatani and Pardeshi 2002). One way of expressing causation in Spanish is through the well-known causative construction (e.g., *El maestro hizo disculparse al alumno* 'the teacher made the student apologize'). From a Construction Grammar (henceforth CxG) perspective, the causative construction can be understood as an abstract schema that links a set of formal and functional features. From a semantic perspective, the first participant, i.e., the causer, acts upon a second participant, i.e., the causee, who is involved in a concatenated event, referred to as the effect or caused event. It is, therefore, a construction composed of two linked events: a causing event and a caused one. From a formal viewpoint, the causative construction is made up of the following elements: [nominal phrase 1 + causative verb + nominal phrase 2 (+ preposition) + infinitive].¹

This overarching causative schema is realized through several micro-constructions which share inherited features, but also present formal and semantic differences. The micro-construction that has received most attention is the one with the causative verb *hacer* ('make') (among others, Cano-Aguilar 1977; Davies 1995; Roegiest and Enghels 2008; Vivanco 2019), defined as the prototypical causative micro-construction due to its

high frequency and semantic versatility (Enghels and Comer 2020). More recently, some attention has been paid to the causative micro-constructions with *dejar* ('let') (Enghels and Roegiest 2014; Maldonado 2007) and with *poner* and *meter* ('put') (Comer 2020; Enghels and Comer 2020; Vivanco 2020). Other Spanish causative verbs, such as *obligar* ('force') or *inducir* ('induce'), with some exceptions (Alfonso Vega 1997), have received little attention in the literature.

The syntactic relationship between the main causative verb and the subordinate infinitive in the causative construction has been debated in the literature. A good number of researchers consider it to be a biclausal structure, assuming that the infinitive and its complements form a phrase subordinate to the causative verb (Alfonso Vega 1997; RAE 2009; Treviño 1994; Vivanco 2019), while others opt for a monoclausal or periphrastic analysis, in which the causative verb functions as an auxiliary² (Comer 2020; Enghels and Comer 2020; Hamplová 1970; Kemmer and Verhagen 1994). In fact, the degree of syntactic incorporation between the causative verb and the infinitive varies between Romance languages (Da Silva 2012) and also within the same language, depending on its evolutionary stage (Davies 1995) or the main causative verb (Comer 2020; Enghels and Comer 2020). For example, Enghels and Comer (2020), in their contrastive study of causative micro-constructions with *hacer* (e.g., *El maestro hizo disculparse al alumno* 'the teacher made the student apologize'), *meter* (e.g., *Metió a su hijo a trabajar* '(s)he put his child to work') and *poner* (e.g., *El presidente puso a debatir a los ministros* 'the president set the ministers to debate'), observe that the variant with *hacer* presents the highest degree of syntactic incorporation of the three.

In the same study, the authors observe that the causative micro-constructions with *meter* and *poner* present a more restricted semantic profile than the one with *hacer*: they are preferably used to designate events of direct causation, express an inchoative aspectual value and are frequently associated with infinitives belonging to the culinary domain (Enghels and Comer 2020, p. 190). As a consequence, the micro-constructions with locative verbs are defined as a particular subschema of the causative construction, in which particular syntactic and semantic features from their locative origin persist. Finally, specific variation among the causative micro-constructions has been related to their stage of grammaticalization. Based on parameters such as the syntactic incorporation between the causative verb and the infinitive and the diversity of semantic contexts they admit, Enghels and Comer (2020) conclude that the micro-construction with *hacer* is more grammaticalized than the micro-constructions with *poner* and with *meter*. Moreover, from a diachronic perspective, the authors observe that the micro-construction with *poner* gradually undergoes an advance in its grammaticalization, unlike the micro-construction with *meter*, whose evolution shows a certain degrammaticalization and a semantic specialization towards increasingly restricted contexts of direct causation and material caused events. Against this background, the present paper provides a diachronic analysis of the understudied causative micro-constructions with the verbs *llevar* 'take' and *traer* 'bring'. From a semantic viewpoint, the micro-construction with *llevar* has been classified within the scope of indirect causation (Da Silva 2004, 2012), which implies that the causer has little control over the caused event, performed by the causee autonomously. Formally, the causative micro-constructions with the verbs *llevar* and *traer* are identified by the preposition *a* before the infinitive, which distinguishes them from other causative micro-constructions that use the verb *hacer*. Thus, the formal structure of these causative micro-constructions is the following: [NP1 + *llevar/traer* + NP2 + *a* + infinitive], as shown in the examples in (1), taken from the *Corpus del Diccionario Histórico* (CDH).

(1) a. El espíritu de intenso lucro que animaba a la economía liberal *llevaba al patrono a imponer* jornadas agotadoras. (Carlos García Oviedo, *Tratado Elemental de Derecho Social*, 1946).

'The spirit of intense profit-making that animated the liberal economy **led** the employer to impose strenuous working days'.

b. El descubrimiento de esta lápida *nos trae por la mano a tratar* de los tesoros depositados en Guarrázar. (Vicente de la Fuente, *Historia eclesiástica de España*, II, 1855–1875).

‘The discovery of this tombstone **leads** us by the hand to deal with the treasures deposited in Guarrázar’.

At first glance, the two micro-constructions appear very similar. Still, in order to better understand the development of the causative construction, this study attempts to describe the formal and functional characteristics of the micro-constructions with *llevar* and *traer*, and to compare their degree of grammaticalization in different stages of Spanish. More specifically, we will try to find out if they experience an advance in their grammaticalization over time, as was the case of the micro-construction with *poner*, or if there is a semantic specialization in increasingly restricted contexts, such as *meter*. On the other hand, as full verbs, *llevar* and *traer* express events in which an agent moves a figure in a controlled way by the translocation of the agent’s body (Talmy 2000, p. 51). Given the semantic relatedness of *llevar* and *traer* as motion verbs, the question arises whether the causative micro-constructions with *llevar* and *traer* are part of a particular subschema, different from that of other causative micro-constructions previously mentioned.

Moreover, we examine the relationship between the historical development of the semantics of *llevar* and *traer* as full verbs and their use in the causative construction. This is in line with the outcomes of previous studies suggesting that the diachrony of grammaticalized verbs can be influenced by changes at the lexical-semantic level, and that both grammaticalized and full lexical meanings of verbs are strongly interrelated in polysemous semantic networks. Garachana Camarero and Hernández Díaz (2020), for instance, show that the emergence of the verbal periphrases <*tener de/que* + infinitive> in Old Spanish occurs by analogy with <*haber de/que* + infinitive> as soon as *tener* had developed its possessive meaning as a full verb, becoming a near-synonym of *haber*. Moreover, the authors show that the periphrases <*haber de/que* + infinitive> have semantically narrowed in the 16th century, as a consequence of the loss of the possessive meaning of *haber* in favor of its existential meaning.

The remainder of the paper is organized as follows. Section 2 focuses on the origin of the causative micro-constructions with *llevar* and *traer*, which leads us to explore the meaning of *llevar* and *traer* as motion verbs and in related metaphorical uses. Section 3 provides more details on the diachronic corpus that serves as an input for the analysis of the causative micro-constructions with *llevar* and *traer*. Section 4 presents the results and describes the outcomes of a productivity analysis of the objects of study, as well as their formal and semantic characteristics. Section 5 further discusses the results and presents the conclusions of this work.

2. From Directed Motion to Causation

In their prototypical meaning, *llevar* and *traer* can be classified as Caused Accompanied Motion verbs (CAM verbs).³ CAM verbs express events in which an entity causes the change in location of a second entity in a particular way: (a) the event causing the movement has a prolonged duration and is simultaneous to the movement itself, and (b) the agent accompanies the displaced entity throughout the motion event, changing his location along with it (Margetts et al. 2022, p. 936). What differentiates *llevar* ‘take’ and *traer* ‘bring’ in present-day Spanish is their component of spatial deixis: *traer* is a coming verb, meaning that it expresses motion towards the deictic center, while *llevar* is a going verb that expresses motion towards places other than the deictic center. However, importantly, this semantic opposition is not constant throughout the diachrony of Spanish. It has indeed been shown that, in Old Spanish, the use of *llevar* and *traer* as motion verbs is not always defined by deictic factors, but that lexical aspect or *Aktionsart* can also play a role (Torres Soler 2023).⁴

In all diachronic stages, the verb *llevar* is used to express motion towards the place occupied by the receiver of the communicative situation (2a) or to a place different from that occupied by both the interlocutors (2b).

(2) a. Otrosi del libro que me enuiaste a mandar que uos **leuasse** sabet sennor que non lo tengo aqui mas yo he enbiado por el e **leuar** uos lo he quando me fuere para uos. (*Colección diplomática de D. Juan Manuel*, 1308).

‘Moreover, about the book that you ordered me to **bring** you, you must know, lord, that I do not have it here, but I have sent someone for it and I will bring it to you when I will come to you’.

b. Se ha fecho y fase en la dicha çibdad de Xeres vna feria, a la qual se **lieua** e descargan muy muchas mercadurías, las quales se **lieuan** de allí a otras partes. (*Tumbo de los Reyes Católicos del concejo de Sevilla*, 1479).

‘It has been done and it is done in the aforesaid town of Jerez a fair to which many goods are **taken** and unloaded, which are **taken** from there to other places’.

In Late Latin, *levare* (the etymon of *llevar*) functioned in one of its meanings as a verb of removal (Cifuentes Honrubia 2008) that expressed movements away from the initial location, without a specific Goal bounding the event at its endpoint. These kinds of movements are referred to as *ablative* or Source-oriented motion events, as opposed to *allative* or Goal-oriented motion events, which consist in moving towards a final location (Kopecka and Vuillermet 2021; Stefanowitsch and Rohde 2004). In Old Spanish (12th–15th centuries), although *llevar* also accepted delimited Goals, it still showed a tendency to express Source-oriented motion events. For instance, in (3), *llevar* is used to refer to displacements bounded by their Source, which is a specific place (*el prado* ‘the meadow’), but not by their endpoint, which it is not specified, because it is not relevant to know where the flowers ended up.

(3) Los omnes e las aves, quantos acaecién, **levavan** de las flores quantas **levar** querién, mas mengua en el prado niguna non facién, por una qe **levavan** tres e quatro nación (Gonzalo de Berceo, *Los Milagros de Nuestra Señora*, 1246–1252).

‘Men and birds, as many as came along, **took** as many flowers as they wanted to **take**, but they did not cause any damage to the meadow; for each one they **took**, three or four were born’.

The verb *traer*, for its part, was used in Old Spanish (12th–15th centuries) to express movements with any deictic orientation, meaning that it could be used to express motion events independently of the place occupied by the interlocutors. However, in the 16th century, the deictic component of its meaning became established, and therefore, in Classical Spanish (16th–17th centuries), it could only be used for motion events towards the area of the speaker or the receiver. In Modern and Contemporary Spanish (18th century–present), the meaning of *traer* further specialized in the expression of motion towards the place of the speaker. The following are examples in which *traer* expresses motion towards the speaker (4a), the receiver (4b) and towards a place different from that occupied by both interlocutors (4c) in Old Spanish.

(4) a. Ydvos para los joyeros e **trahedme** composturas. (*La historia de la donzella Teodor*, 1250).

‘Go to the beauticians and **bring** me cosmetics’.

b. Vi vuestra carta en que me enviastes dezir que, por vos fazer yo bien e merçed, tove por bien e mande que ningun omne non **troxiesse** a vuestra villa por dos annos vino de fuera parte. (*Documentos de Alfonso X dirigidos al Reino de León*, 1271).

‘I saw your letter in which you told me that, in order to favour you, I decided and ordered that nobody would **bring** wine from outside to your village for two years’.

c. Vuestros subditos, [. . .], los quales deziys tienen licencia de nuestro Sancto Pedro de negociar e **traher** madera, pez e otras mercadurias a las tierras de los turcos e de los egipcianos. (‘Fernando al cardenal maestre de Rodas [. . .]’, *Documentos sobre relaciones internacionales de los Reyes Católicos*, II, 1496).

‘Your subjects, [. . .], who you say that they have licence from our Holy Father to negotiate and **take** wood, fish and other goods to the lands of the Turks and the Egyptians’.

As regards its Aktionsart, *traer* in Old Spanish functioned as an allative or Goal-oriented motion verb that specialized in the expression of motion events with a specific Goal bounding the event at its endpoint. For instance, in (5), *traer* expresses a motion event that is bounded by the Goal *al dicho monasterio* ('to the said monastery').

(5) Ellos siendo llamados algunas veses por los dichos abad e convento sus señores, o por su mandado, que no quisieron **traer** el pan e la sal con sus bestias al dicho monesterio (*Colección documental de Alfonso XI, 1347*)

'They, being sometimes called upon by the said lords abbot and members of the monastery, or by their command, did not want to bring the bread and salt with their beasts to the said monastery'.

Therefore, in Old Spanish, *llevar* and *traer* competed in the expression of motion towards places not occupied by either of the interlocutors, and the choice of one or the other was influenced by their lexical aspect. In these contexts, Source-oriented movements without a concrete Goal are mostly expressed by *llevar*, while *traer* is used preferably in contexts where a specific Goal bounds the motion event. In example (6), *llevar* and *traer* express motion to places not occupied by the interlocutors. *Traer* is employed when there is specific Goal (*a la dicha feria* 'to the said fair') bounding the event at its endpoint, while *llevar* is chosen when the Source is a specified place, namely the same fair, and the Goal is not concretized.

(6) E a todo lo que consigo **truxieren** a la dicha feria e toiueren en ella e **lleuaren** della (*Tumbo de los Reyes Católicos del concejo de Sevilla, 1479*).

'And everything that they **carry** with them to the said fair and that they have in there and **take** away from there'.

The aspectual differences of *llevar* and *traer* in Old Spanish were progressively weakened throughout the Middle Ages in favour of their incipient deictic value, until *llevar* became the only verb capable of expressing motion to a place different from the deictic center in the 16th century. In short, three stages can be distinguished in the history of *llevar* and *traer* as motion verbs: (a) in Old Spanish (12th to 15th centuries), their deictic opposition was not yet fully consolidated, so their distribution depended on both incipient deictic factors and Aktionsart differences, (b) in Classical Spanish (16th to 17th centuries), their opposition was already fully deictic and the deictic range of *traer* comprised both the 1st and 2nd person, and (c) in Modern Spanish (18th to 20th centuries), their opposition was still fully deictic and the deictic range of *traer* was restricted to the first person. Parallel to their meaning as motion verbs, *llevar* and *traer* developed a series of metaphorical extensions. In concrete, we hypothesize that the causative micro-constructions under study derive from metaphorical expressions that indicate the causation of a process as a result of which the direct object gets involved in a new event. This event, which can be either static or dynamic, is mentioned by means of a nominal expression preceded by a preposition (usually *a*). It is metaphorically conceptualized as an abstract place reached by the direct object, thus as the Goal of a metaphorical motion event. The underlying conceptual metaphor, as formulated by Lakoff and Johnson (1999), is CAUSATION IS FORCED MOVEMENT. On the one hand, in the source domain of caused motion, we typically find an agent (NP1) that exerts a physical force to displace a figure (NP2) that receives the driving force and consequently moves to a new location (<a + NP3>). An example is *María llevó el jarrón al salón* ('María took the vase to the hall'). On the other hand, in the target domain of causation, we find a causer (NP1), animate or not, that causes a causee (NP2) to be involved in an event (<a + NP3>) (7).

(7) a. El uno de tus amigos es aquel que te tuelle del mal y te **lieva** al bien (*Libro de los buenos proverbios que dijeron los filósofos y sabios antiguos, 1250*)

'Your friend is the one who takes you out of evil and **leads** you to good'.

b. Quando la sacra excelencia nuestra yglesia començo vnos **truxo** con clemencia y otros a la penitencia (Pedro de Salinas, *Poemas, 1533*).

‘When the sacred excellence of our Church began, it **led** some with mercy and others to penance’.

In examples (7a–b), the nominal expressions preceded by *a* no longer represent the Goal of a displacement in space, but an event in which the direct object gets involved. In example (7a), *al bien* (‘to good’) is a state, whereas in (7b), *a la penitencia* (‘to penance’) is a dynamic event.

The causative micro-constructions with *llevar* and *traer* originate from the aforementioned metaphorical expressions, when the event in which the direct object gets involved is no longer expressed by a nominal expression, but through an infinitive. The close conceptual relationship between these metaphorical expressions and the infinitive construction is corroborated by examples such as (8), which is the oldest documented example of the causative micro-construction with *llevar*.

(8) Sean vuestras cobdicias **levadas** a aver buena fama; non las **levedes** a las malicias nin a las feas cosas. (*Bocados de oro*, 1250).⁵

‘Let your ambitions be **led** to have a good reputation; do not **lead** them to evil and ugly things’.

Example (8) illustrates a caused event expressed by the infinitive and its complement *aver buena fama* (‘to have a good reputation’), which is opposed in the following sentence to the abstract locative complement *a las malicias nin a las feas cosas* ‘to evil and ugly things’. Both sentences imply that the direct object is caused to get involved in an event, but only in the second this event is transparently conceptualized as an abstract Goal. In short, the origin of the causative micro-constructions with *llevar* and *traer* can be summarized as follows: (a) ‘to lead (*llevar/traer*) to a concrete place’ >, (b) ‘to lead (*llevar/traer*) to an abstract place’ >, (c) ‘to lead (*llevar/traer*) to an event’.

In light of the semantic changes experienced by *llevar* and *traer* as full lexical verbs throughout the history of Spanish, the following sections explore whether they had an impact on the use of *llevar* and *traer* in the causative construction. More specifically, it is examined whether the evolving semantic differences between the verbs had an impact on the overall evolving productivity and varying degrees of grammaticalization of the causative micro-constructions under study.

3. Methodology

3.1. Data Collection

For the analysis of the causative micro-constructions, relevant data were extracted from the *Corpus del Diccionario histórico de la lengua española* (CDH), an extensive diachronic corpus of Spanish elaborated by the Real Academia Española.⁶ More specifically, we used two of the three query layers of the CDH, namely the *nuclear CDH* layer, consisting of texts from all periods up to the year 2000, characterized by their strict philological rigor, and the *S. XII-1975* layer, which allowed us to reach a broader view of the construction, thanks to the large number of historical documents included. However, the *1975–2000* layer was not used because the number of texts that it contains from a 25-year period is too large for the macro-diachronic interests of this paper. Since the focus is not on diatopic variation, only data from Spain were considered.

We downloaded all occurrences of the lemmas *llevar* and *traer* (in any of their forms) followed on the right by the preposition *a* and an infinitive. We included cases in which the causative verb and the <*a* + INF> group are contiguous as well as cases in which they are separated by up to 3 intercalated linguistic elements.⁷ In a next phase, the data were manually checked in order to exclude all non-relevant cases from the dataset. As a result, we obtained a corpus of 1188 tokens of the causative micro-construction with *llevar* and 218 tokens with *traer*.

3.2. Parameters of Analysis

To begin with, the internal dynamics of the causative construction is explored through a productivity analysis (Section 4.1). The productivity analysis centers around token frequency, which measures the general frequency of the micro-constructions, and type frequency, which accounts for their semantic spread and density (Barðdal 2008). It is particularly interesting to investigate (a) to what extent both micro-constructions with *llevar* and *traer* have competed in the different stages of Spanish, and (b) whether their degree of productivity was affected by the semantic changes experienced by *llevar* and *traer* as full lexical verbs.

Second, in order to account for the degree of grammaticalization of the causative micro-constructions with *llevar* and *traer* in the different stages of Spanish, the dataset has been annotated according to a series of well-defined formal and semantic parameters. In concrete, four formal parameters are analyzed with a focus on the fixation of the construction and its degree of unithood. First, in order to study the degree of formal fixation, the type of preposition introducing the infinitive complement will be examined. In fact, in addition to the preposition *a* ('to'), structural variants with the preposition *en* ('in') and without prepositions have been attested in the corpus. This observation raises the question whether a process of morphosyntactic specialization (in terms of Hopper 1991) or obligatoriness (in terms of Lehmann 2015) has occurred, which would indicate further progress in the grammaticalization process (Section 4.2.1).

Next, the notion of unithood refers to the conceptualization of the causative situation as a single action, rather than two autonomous events. A high degree of unithood relates to a higher degree of grammaticalization of the construction (Enghels and Comer 2018). The degree of unithood can be observed through the degree of incorporation or syntagmatic linkage between the causative verb and the infinitive. This criterion is operationalized by looking into the presence or absence of intercalated linguistic elements between the causative verb and the infinitive, in concrete the causee (Section 4.2.2) and other adjuncts (Section 4.2.3). A fourth criterion that provides further insight into the degree of unithood of the construction is the degree of association between the causative verb and the infinitive (Comer 2020; Torres Cacoullous 2000). While the combination of the causative verb with a single infinitive indicates a high degree of association between the two, the presence of multiple infinitives indicates a lower degree of association, and consequently, a lower degree of grammaticalization (Section 4.2.4).

Finally, in order to comprehend the functional development of the causative construction, two semantic parameters are examined. First, in grammaticalization processes, the elements of the construction undergo a loss of semantic integrity (Lehmann 2015, pp. 126–28). In the case of verbal grammaticalization, this process takes the form of a relaxation of the semantic constraints operating in the selection of the verbal complements (Lamiroy 1994, p. 35). Moreover, from a CxG perspective, it has been observed that, as a constructional schema develops and becomes more abstract the construction is expected to expand semantically to new domains, thus increasing the repertoire of linguistic units with which it can be combined (Barðdal 2008; Enghels and Comer 2018). In order to investigate if the causative micro-constructions with *llevar* and *traer* enhanced their semantic scope over time, the animacy of both the causer and the causee (Section 4.3.1) and the semantic types of the infinitives (Section 4.3.2) are examined.

4. Results

4.1. General Productivity Measures

Linguistic units involved in a grammaticalization process tend to experience an increase in their token frequency (Traugott and Heine 1991, p. 9). Similarly, diachronic CxG studies point out that constructionalization and constructional changes coincide with a rise in token frequency (Traugott and Trousdale 2013). Consequently, an increasing token frequency of the micro-constructions with *llevar* and *traer* might be an indicator of their grammaticalization and a sign of the further development of the causative construction

as a schema. While absolute token frequency counts the total number of occurrences of the micro-constructions in the corpus, the measure of relative token frequency shows their standardized frequency (e.g., per million words).

Next, type frequency accounts for the number of different lexical items that can fill an open slot in the construction (Barödal 2008), in our case, the infinitive slot. When a construction is involved in a grammaticalization process, it is expected to expand to new contexts, to increase its productivity and thus to enlarge its collocational range (Traugott and Trousdale 2013). Consequently, an increase in the type frequency of the causative micro-constructions with *llevar* and *traer* should be interpreted as a sign of the expansion of the construction to new semantic domains. Again, it is informative to take into account both the absolute type frequency, which indicates the total number of different infinitives, and the relative type frequency, which measures the number of infinitives documented per million words. Table 1 shows the token frequency and the type frequency of the causative micro-constructions with *llevar* and *traer*, both in absolute and relative numbers, in the period between the 13th and 20th century.

Table 1. Token and type frequencies of the causative micro-constructions with *llevar* and *traer*.

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------------|--------------------------|------|------|------|------|------|------|-------|
| | Absolute token frequency | | | | | | | |
| <i>Llevar</i> | 1 | 1 | 9 | 48 | 49 | 19 | 254 | 807 |
| <i>Traer</i> | 6 | 13 | 59 | 86 | 37 | 2 | 14 | 1 |
| Total | 7 | 14 | 68 | 134 | 86 | 21 | 268 | 808 |
| | Relative token frequency | | | | | | | |
| <i>Llevar</i> | 0.12 | 0.13 | 0.34 | 0.93 | 1.38 | 1.54 | 6.42 | 15.41 |
| <i>Traer</i> | 0.71 | 1.66 | 2.24 | 1.67 | 1.07 | 0.16 | 0.35 | 0.02 |
| Total | 0.83 | 1.79 | 2.58 | 2.6 | 2.45 | 1.7 | 6.77 | 15.43 |
| | Absolute type frequency | | | | | | | |
| <i>Llevar</i> | 1 | 1 | 6 | 34 | 38 | 18 | 173 | 392 |
| <i>Traer</i> | 5 | 8 | 35 | 43 | 26 | 2 | 12 | 1 |
| Total | 6 | 9 | 41 | 77 | 64 | 20 | 185 | 392 |
| | Relative type frequency | | | | | | | |
| <i>Llevar</i> | 0.12 | 0.13 | 0.23 | 0.66 | 1.07 | 1.45 | 4.37 | 7.49 |
| <i>Traer</i> | 0.59 | 1.02 | 1.33 | 0.83 | 0.73 | 0.16 | 0.30 | 0.02 |
| Total | 0.71 | 1.15 | 1.56 | 1.49 | 1.8 | 1.61 | 4.67 | 7.51 |

If we focus on the productivity of both causative micro-constructions together (see the numbers starting with ‘Total’), it is observed that type and token frequencies slowly increase between the 13th and 15th centuries. This can be interpreted as a sign of the development of a subschema of the causative-construction with CAM verbs and its increasing grammaticalization. Between the 15th and 18th centuries, its productivity remained relatively stable, with even a small decrease in the 18th century.⁸ From the 19th century on, the productivity of the subschema increased strongly. Thus, the productivity profile of the subschema follows the typical S-shaped curve with a slow beginning followed by a rapid expansion (Nevalainen 2015). Considering this rise in productivity, it is to be expected that the subschema has undergone formal changes over time that reveal an increasing grammaticalization pattern.

Next, when comparing both micro-constructions, the results of the productivity analysis show that the causative micro-construction with *traer* was by far the most productive one in Old Spanish. Its relative type and token frequencies reach its maximum productivity in the 15th century (resp. 2.24 and 1.33), after which it starts to lose productivity until it becomes marginal from the 18th century onwards (e.g., with a relative token frequency of 0.02 in the 20th c.). On the contrary, the micro-construction with *llevar* displays an uninterrupted increase in its type and token frequencies throughout the history of Spanish, surpassing the

micro-construction with *traer* in the 17th century and experiencing a particularly marked increase in productivity in the 19th and 20th centuries.

In the light of these data, we can distinguish three phases in the internal development of the subschema, which coincide with the well-known periodization of Spanish in three stages (Cano-Aguilar 1988): (a) a first phase in which the micro-construction with *traer* was dominant, which corresponds to the Old Spanish period (13th to 15th centuries), (b) a phase of competition between both micro-constructions, during the Classical Spanish period (16th and 17th centuries) and (c) a last phase in which the micro-construction with *llevar* is clearly dominant, which coincides with the Modern Spanish period (18th to 20th centuries). Thus, the main changes in the internal development of the subschema coincide with two moments of major transformations in the linguistic system of Spanish. The beginning of the decrease in productivity of *traer* in favor of *llevar* takes place during the transition between Old and Classical Spanish, while the total decline of the micro-construction with *traer* occurs in the “Early Modern Spanish” period (Octavio de Toledo Huerta 2016), which spans from the end of the 17th century to the beginning of the 19th century.

The predominance of *traer* in Old Spanish and the later development of the micro-construction with *llevar* can be explained by the Aktionsart characteristics of *llevar* and *traer* as motion verbs (see Section 2). We already know that the subschema of the causative construction with CAM verbs has its origin in metaphorical expressions in which an event is conceptualized as the Goal of a displacement. Therefore, for a motion verb to enter into the subschema, it must be able to express displacements towards a syntactically realized Goal. The tendency of *llevar* to express Source-oriented movements without a Goal bounding the event hindered its incorporation into the subschema in Old Spanish. In contrast, the tendency of *traer* to express Goal-oriented motion events facilitated the productivity of the causative micro-construction with *traer* in the early period, compared to *llevar*.

From the 16th century on, the loss of productivity of the micro-construction with *traer* and the expansion of the micro-construction with *llevar* can be explained by the consolidation of the deictic opposition of *llevar* and *traer* as motion verbs, and the subsequent loss of their original aspectual properties. In the 16th century, *traer* has become entirely a coming verb, hence the marked term of the verb pair.⁹ Since there is no connection between the deictic center and the caused event, the micro-construction with *traer* became less semantically transparent, and as a consequence, its productivity decreased. On the contrary, once *llevar* lost its tendency to express Source-oriented motion events, it became more suitable to be used in the causative construction. The beginning of the last phase, in which *traer* is residual and *llevar* is clearly predominant (18th to 20th centuries), coincides with a further restriction of the meaning of *traer* as a deictic motion verb, so that it can only express movements towards the place occupied by the speaker.

4.2. Formal Features

The increase in productivity of the potential subschema with CAM verbs, led by the micro-construction with *llevar*, can be an indicator of its ongoing grammaticalization. It is possible, therefore, that the micro-constructions under study, especially the one with *llevar*, have undergone formal changes typical of grammaticalization processes. This subsection examines four formal parameters. The first one concerns the fixation of the micro-constructions based on a variational analysis of the prepositional slot (Section 4.2.1). The three other parameters, namely, the position of causee (Section 4.2.2), the degree of adjacency (Section 4.2.3) and association (Section 4.2.4), measure and compare their degree of unithood.

4.2.1. Fixation of the Prepositional Slot

As mentioned in Section 1, the causative micro-constructions with *llevar* and *traer* typically present the following formal structure: [NP1 + *llevar/traer* + NP2 + *a* + infinitive]. However, the corpus analysis shows that, in former stages, other structural variants existed without prepositions, as in (9a–b), or with the preposition *en* (‘in’) instead of *a* (‘to’) (9c).¹⁰

(9) a. Deseaba que ya que mi buena suerte *me había llevado* gozar de su alegre vista, me concediese lograr el llamarme suya. (Andrés Sanz del Castillo, *La mojiganga del gusto*, 1641).

‘I wished that, since my good fortune had **led** me to enjoy his joyful sight, it would grant me to call me his’.

b. El juego, que por sí mismo y de su naturaleza, *trae* recrear el ánimo. (Francisco de Luque Fajardo, *Fiel desengaño contra la ociosidad y los juegos*, 1603).

‘Gambling, which by itself and by its nature, **leads** to entertain the spirit’.

c. La composición del hombre *nos deve traer en alabar y servir y amar* al artífice, que es Dios. (Pedro Mejía, *Silva de varia lección*, 1540–1550).

‘The composition of the human being should **lead** us to praise and serve and love the artificer, who is God’.

Since the preposition is a fixed component of the construction, the reduction in the paradigmatic variability of this slot could be interpreted as an indication of the grammaticalization of the pattern (Lehmann 2015). Therefore, all the occurrences in the CDH responding to the following formal structures were collected: [NP1 + *llevar/traer* + NP2 + *en* + infinitive] and [NP1 + *llevar/traer* + NP2 + infinitive].¹¹ Table 2 presents a historical overview of the absolute frequencies of the micro-constructions under study and the variants without prepositions and with the preposition *en*.

Table 2. Constructional variants with the preposition *en* and without prepositions.

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----------------------------------|----|----|----|----|----|----|-----|-----|
| <i>Llevar</i> + < <i>a</i> + INF> | 1 | 1 | 9 | 48 | 49 | 19 | 254 | 807 |
| <i>Llevar</i> + <∅ + INF> | - | - | 1 | - | 1 | - | - | - |
| <i>Traer</i> + < <i>a</i> + INF> | 6 | 13 | 59 | 86 | 37 | 2 | 14 | 1 |
| <i>Traer</i> + < <i>en</i> + INF> | - | 1 | 1 | 1 | - | - | - | - |
| <i>Traer</i> + <∅ + INF> | - | 2 | 14 | 1 | 1 | - | - | - |

The results of the Table 2 show that micro-constructions with the preposition *a* were most frequent in all periods, and that the variant with the preposition *en* has always been exceptional. The causative micro-construction with *traer* showed some paradigmatic variation between the preposition *a* and absence of a preposition, especially in the 15th century. On the contrary, the causative micro-construction with *llevar* has presented a higher degree of formal fixation all over time, as only two occurrences without prepositions could be retrieved. This indicates that the micro-construction with *llevar* presents lower paradigmatic variability than the one with *traer*. However, from the 18th century onwards, the whole subschema became fixed with the preposition *a*, what is a sign of its grammaticalization (Lehmann 2015).

4.2.2. Position of the Causee

In grammaticalization processes, the syntagmatic linkage between the elements of a construction is gradually reinforced, which is known as syntactic incorporation (Correia Saavedra 2021; Lehmann 2015). In verbal grammaticalization, it is thus expected that the presence of lexical items interposed between the auxiliary verb and the non-finite verb will become less frequent over time (Heine 1993; Rodríguez Molina 2010; Torres Cacoullous 2000). Regarding the causative construction, the formal variable that has probably most attracted the attention of scholars is the position of the causee (NP2) when lexically expressed (Engiels and Comer 2020; Engiels and Roegiest 2014; Shibatani and Pardeshi 2002; Da Silva 2004). More concretely, the placement of the causee between the causative verb and the infinitive, which has been called VOV structure, implies a lower degree of incorporation, and thus grammaticalization, between the two verbs (10).

(10) a. El despecho *llevaba a Pablo a hacer* alarde de una indiferencia despreciativa (Fernán Caballero. (Cecilia Böhl de Faber), *Clemencia*, 1852).

‘The spite **led** Pablo to display a disdainful indifference’.

b. Onde devedes saver que quando el peccado non es purgado por penitencia, por la su grand pesadumbre *trahe al peccador a caher* en otros. (*Un sermonario castellano medieval*, 15th century).

‘Therefore, you must know that when the sin is not purged by penance, for its big weight it **leads** the sinner to fall into others’

On the contrary, the postposition of the causee, which has been called VV structure, is interpreted as a sign of more incorporation and unithood, hence a more advanced stage of grammaticalization of the construction (11).

(11) a. Las observaciones anteriores *han llevado a afirmar a los escrituristas* que el pueblo hebreo y, por lo tanto, la Biblia, tiene un cierto sentido lineal del tiempo. (Luis Maldonado, *La plegaria eucarística. Estudio de teología bíblica y litúrgica sobre la misa*, 1967).

‘The above observations have **led** writers to assert that the Hebrew people, and therefore the Bible, have a certain linear sense of time’.

b. el demonio tienta experimentalmente si podrá *traer a pecar a los hombres*. (Juan de Pineda, *Diálogos familiares de la agricultura cristiana*, 1589).

‘The Devil tries experimentally if he will be able to **lead** people to commit sins’.

According to Da Silva (2004), the two syntactic schemes—VOV and VV—profile different elements of the event: the VOV scheme profiles the causee, presented as a more autonomous entity, while the VV scheme profiles the caused event as a whole. Thus, the VOV structure is related to indirect causation events, in which the causee acts autonomously, while the VV structure is linked to direct causation, so the causer has more control over the caused event (Section 4.3.1 further details the concepts of direct and indirect causation). Considering that the micro-constructions under study have been linked to indirect causation, the VOV structure is expected to predominate. However, over the centuries, further grammaticalization of the construction would entail an increasing number of examples with the VV structure. Table 3 presents quantitative data on the position of the nominal causee per century.¹²

Table 3. Position of the NP2 in the micro-constructions with *llevar* and *traer*.

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------------------|----------|----------|----------|----------|---------|----------|----------|-----------|
| <i>Llevar</i> VOV | - | - | 3 (100%) | 5 (100%) | 1 (33%) | 2 (100%) | 23 (96%) | 111 (94%) |
| <i>Llevar</i> VV | - | - | - | - | 2 (67%) | - | 1 (4%) | 7 (6%) |
| Total <i>llevar</i> | - | - | 3 | 5 | 3 | 2 | 24 | 118 |
| <i>Traer</i> VOV | 1 (100%) | 3 (100%) | 11 (92%) | 8 (80%) | 3 (75%) | - | 2 (100%) | - |
| <i>Traer</i> VV | - | - | 1 (8%) | 2 (20%) | 1 (25%) | - | - | - |
| Total <i>traer</i> | 1 | 3 | 12 | 10 | 4 | - | 2 | - |

Table 3 shows that the causative micro-constructions with *llevar* and with *traer* are preferably expressed through the VOV structure in all epochs. The only exception is the micro-construction with *llevar* in the 17th century, although very few examples are examined.

The overall image points to a constant low degree of incorporation and unithood in both CAM-based micro-constructions. This is quite remarkable given that other causative subschemas show a diachronic tendency towards increasing postposed causees. For example, the proportion of examples with the postposed causees in the micro-construction with *hacer* increases from 50% in Old Spanish to 92% in Modern Spanish (Davies 1995). Similarly, in the micro-construction with *poner*, the postposition of the causee represents 25% of examples in the 13th century and ascends to 60.8% in the 21st century (Engels and Comer 2020). This suggests that the potential subschema with CAM verb is a rather fixed pattern with very little syntactic variation across time.

4.2.3. Position of Adjuncts

In addition to the causee, adjunct complements (including adverbs and prepositional phrases) allow high mobility in the construction. Consequently, their position can also shed light on the degree of incorporation between the causative verb and the infinitive, and hence on the degree of unithood of the construction (Enghels and Comer 2018; Rodríguez Molina 2010; Torres Cacoullós 2000). Adjuncts can be placed between the causative verb and the infinitive (12), or before or after these constituents (13).

(12) a. No será éste un método que *lleve simplemente a resolver* los problemas que las cosas plantean. (Xavier Zubiri, *Naturaleza, Historia, Dios*, 1932–1944).

‘This will not be a method that **leads** simply to solving the problems that things pose’.

b. La congoja, y deseo la *traía muchas veces* á desfallecer y desmayarse. (Fray Luis de León, *Exposición del Cantar de los Cantares*, 1561).

‘The distress and desire **led** her many times to give out and faint’.

(13) a. Lo que acabamos de decir *nos lleva a tratar ahora* de la constitución general de nuestro globo. (S. Alvarado, *Ciencias Naturales*, 1957–1974).

‘What we have just said leads us to address now the general constitution of our globe’.

b. Suplica a Dios omnipotente [. . .], que por su alto juicio *le ha traído a suceder* en esta monarquía. (Luis Cabrera de Córdoba, *Historia de Felipe II, rey de España*, 1619).

‘He begs almighty God [. . .], which by his high judgement has **led** him to succeed in this monarchy’.

Thus, it is expected that a higher degree of grammaticalization of the micro-constructions with *llevar* and *traer* coincides with less adjuncts to be found between the causative verb and the infinitive. Table 4 shows the quantitative results of interpolation of adjuncts.¹³

Table 4. Interpolation of adjuncts between the causative verb and the infinitive.

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------------------------------|----|---------|----------|----------|----------|---------|----------|----------|
| <i>Llevar</i> interpolation | - | - | 2 (67%) | 4 (27%) | 4 (27%) | 4 (36%) | 26 (36%) | 70 (45%) |
| <i>Llevar</i> non-interpolation | - | - | 1 (33%) | 11 (73%) | 11 (73%) | 7 (64%) | 46 (64%) | 87 (55%) |
| Total <i>llevar</i> | - | - | 3 | 15 | 15 | 11 | 72 | 157 |
| <i>Traer</i> interpolation | - | 1 (50%) | 1 (8%) | 5 (28%) | 3 (25%) | - | 1 (17%) | - |
| <i>Traer</i> non-interpolation | - | 1 (50%) | 11 (92%) | 13 (72%) | 9 (75%) | - | 5 (83%) | - |
| Total <i>traer</i> | - | 2 | 12 | 18 | 12 | - | 6 | - |

The data show that, unlike the causee (NP2),¹⁴ adjuncts are most frequently anteposed to the auxiliary verb or postposed to the infinitive, rather than interposed between the causative verb and the infinitive in both micro-constructions. Except for the 14th century, where there are only two examples with adjuncts, the micro-construction with *traer* shows a clear predominance of the pattern without interpolation and a relatively low proportion of intercalated adjuncts, ranging between 8% and 28%. Regarding the micro-construction with *llevar*, the pattern without interpolation also predominates, but to a lesser extent, since the number of interposed adjuncts ranges between 27% and 45%. An exception is the 15th century, in which the higher proportion of examples with interpolation is due to the small number of relevant cases observed.

However, a closer look at the data shows that the proportion of interposed adjuncts is very similar in both constructions in the 16th and 17th centuries (between 25% and 28%), but after the 18th century, it increases in the micro-construction with *llevar*, reaching its peak in the 20th century (45%). This suggests that, after a period in which both micro-constructions showed a similar degree of syntactic incorporation, unexpectedly, the micro-construction with *llevar* gradually became less incorporated. From a more general perspective, the data

show that the potential subschema with CAM verbs does not seem to reach a higher degree of syntactic incorporation over time.

4.2.4. Degree of Association

The last formal parameter focuses on the combination of the causative verb with more than one (14) or one (15) infinitive.

(14) a. Una poderosa inclinacion *me lleva á respetarlo, y á prestarle* en todas ocasiones testimonios de veneracion. (Fray Francisco Alvarado, *Cartas críticas del Filósofo Rancio*, 1811).

‘A powerful inclination **leads** me to respect him, and to render him on all occasions testimonies of veneration’.

b. Este abominable vicio que casi del todo *trae a aborrescer y menospreciar* a Dios. (Jaime Montañés, *Espejo de bien vivir y para ayudar a bien morir*, 1573–1577).

‘This abominable vice that almost entirely **leads** one to hate and despise God’.

(15) a. El mismo estado de debilidad en que se siente, y que *le lleva á buscar* la fuerza de que carece como individuo aislado. (Casildo Ascárate y Fernández, *Insectos y criptógamas que invaden los cultivos en España*, 1893).

‘The same state of weakness in which he feels, and that leads him to seek the strength that he lacks as an isolated individual’.

b. La fortuna, [. . .], me *ha traído a referir* adversidades. (Francisco Manuel de Melo, *Historia de los movimientos, separación y guerra de Cataluña*, 1645).

‘Fortune, [. . .], has **brought** me to report adversities.’

This parameter is based on the premise that the association of an auxiliary with a single non-finite verb is a sign of a higher degree of unithood of the construction and, consequently, more advanced grammaticalization. On the contrary, the combination of an auxiliary with several non-finite verbs indicates that the auxiliary has a broader structural scope, which is considered as a potential sign of a lower degree of grammaticalization (Comer 2020; Torres Cacoullos 2000). Based on this parameter, Comer (2020) observes that, in present-day Spanish, the causative micro-construction with *meter* combines more frequently with multiple infinitives than the more grammaticalized variant with *poner*. If the causative micro-constructions with *llevar* and *traer* are undergoing a grammaticalization process, a decrease in the frequency of examples with multiple infinitives is expected. Table 5 shows the results.

Table 5. Degree of association between the causative verb and the infinitive.

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----------------------------|----------|-----------|----------|----------|----------|----------|-----------|-----------|
| <i>Llevar</i> one INF | 1 (100%) | 1 (100%) | 9 (100%) | 43 (90%) | 47 (96%) | 14 (74%) | 229 (90%) | 749 (93%) |
| <i>Llevar</i> multiple INFs | - | - | - | 5 (10%) | 2 (4%) | 5 (26%) | 25 (10%) | 58 (7%) |
| Total <i>llevar</i> | 1 | 1 | 9 | 48 | 49 | 19 | 254 | 807 |
| <i>Traer</i> one INF | 6 (100%) | 13 (100%) | 56 (95%) | 78 (91%) | 34 (92%) | 2 (100%) | 13 (93%) | 1 (100%) |
| <i>Traer</i> multiple INFs | - | - | 3 (5%) | 8 (9%) | 3 (8%) | - | 1 (7%) | - |
| Total <i>traer</i> | 6 | 13 | 59 | 86 | 37 | 2 | 14 | 1 |

The data show that both micro-constructions are mostly combined with a single infinitive. In the case of *llevar*, between the 13th and 15th centuries, no cases with multiple infinitives are documented, which may be due to the low number of tokens overall. Between the 16th and 20th centuries, the proportion of examples with multiple infinitives ranges between 4% and 10%, except for the 18th century, what may also be partially explained by the relative data scarcity (see note 7). As far as *traer* is concerned, cases with multiple infinitives are documented with a frequency that ranges between 5% and 9% between the 15th and 17th centuries and in the 19th century. In the other centuries, no examples with

multiple infinitives are found, due to the low number of occurrences. These data suggest that the micro-constructions under study present a comparable and relatively stable degree of association between the causative verb and the infinitive.

To recap, the analysis of a set of formal features has shown that both CAM micro-constructions have evolved in a rather similar way. Regarding its degree of grammaticalization, the causative subschema with CAM verbs is characterized by a first stage of variation in the prepositional slot followed by the fixation of the preposition *a* from the 18th century on. However, the three remaining parameters indicate that there is no increase in the degree of unithood of the construction over time, hence no further grammaticalization. More concretely, there is (a) a constant predominance of the VOV pattern, (b) a relatively stable degree of association between the causative verb and the infinitive and (c) a growing proportion of interposed adjuncts suggesting even a decreasing degree of syntactic incorporation.

The observation that the subschema has not evolved towards greater unithood over time could be due to a hypothetical semantic specialization in the expression of indirect causation (see Section 4.3.1). Constructions that specialize in indirect causation are typically associated with the VOV pattern (Da Silva 2004) and have a lower degree of grammaticalization than those expressing direct causation events (Shibatani and Pardeshi 2002). In the following section, the functional analysis of the micro-constructions will reveal that the lack of an increase in the unithood of the subschema is indeed motivated by semantic factors.

4.3. Semantic Features

This section presents the results of a functional analysis of the causative micro-constructions with *llevar* and *traer* based on two parameters: the animacy of the participants (Section 4.3.1) and the semantics of the infinitive (Section 4.3.2). It explores whether the variants expand to new semantic domains over time, what is expected when a more abstract subschema develops or whether they have specialized in the expression of a specific type of causative event.

4.3.1. Animacy of the Causer and the Causee

The most widespread semantic classification of causative events distinguishes between direct and indirect causation. In an event of direct causation, the causer physically manipulates the causee to perform an action, so that the causee has no autonomy or control over the caused event, but the causer does. In a process of indirect causation, on the contrary, the causer exerts a more restricted influence on the causee and, as a result, the causee autonomously performs an action, which may take place at a different time and place (Shibatani and Pardeshi 2002; Lavale 2013). According to Enghels and Comer (2020), prototypical direct causation has an animate causer and an inanimate causee, as in example (16a), whereas prototypical indirect causation involves an inanimate causer and an animate causee, as in (16b).

(16) a. Bien asy como *el libre escriuano trae la pluma a escreuir* las razones que el quiere. (Mose Arragel de Guadalquivar, *Traducción y glosas de la Biblia de Alba*, 1422–1433).

‘Right so as the free scribe **makes** the quill write the reasons that he wants’.

b. *Mi triunfo primero me llevo á buscarlos* continuos, y á conseguirlos tambien (Antonio Alcalá Galiano, *Memorias*, 1847–1849).

‘My first triumph **led me** to look for them continuously, and to get them too’.

In examples such as (16a), the causer (in this case, *el libre escriuano* ‘the free scribe’) is an animate, agentive being that physically manipulates the causee (*la pluma* ‘the quill’), an inanimate entity lacking all control and autonomy. Consequently, the caused event (*escribir* ‘write’) necessarily occurs simultaneously with the causing event (the action of the scribe on the quill). In contrast, in (16b), the causer is an inanimate entity (in this case, a dynamic state of affairs, *mi triunfo primero* ‘my first triumph’) that influences the causee (*me* ‘me’) in such a way that the latter, an animate and agentive being, gets involved in another action.

Thus, the causing event does not necessarily occur in the same moment or place as the caused event. In fact, in (14b), it is inferred that the causing event (the first triumph) occurs first and that, subsequently, the caused event is carried out (*buscarlos continuos [los triunfos]* ‘look for them continuously [the triumphs]’). Thus, the examples with animate causer and inanimate causee, as well as the examples with inanimate causer and animate causee, clearly correspond to the categories of direct causation and indirect causation, respectively, and represent prototypical cases of each type. Other types of events may lie at intermediate points on the continuum between these two prototypes. Several attempts have been made to characterize and classify the intermediate cases between prototypical direct and indirect causation (e.g., Comer 2020, pp. 486–91; Shibatani and Pardeshi 2002). Besides the animacy of the participants, other factors, such as dynamicity or control, are also involved (Engels and Comer 2020). For instance, the combination of two animate participants mostly results in a situation where the causer induces the causee to perform an action by psychological means (Guilquin 2010), as in example (17a). However, there are also some cases in which the causer acts physically on the causee to force him or her do something, as in example (17b), extracted from Comer (2020, p. 488).

(17) a. *Alcina*, de quien dize Orlando, que por engaño **traya los hombres** a gozar de sus regalos. (Juan de Luna, *Diálogos familiares en lengua española*, 1619).

‘*Alcina*, of whom Orlando says that she deceitfully **led men** to enjoy her gifts’.

b. La levanta del asiento, engancha sus manos entre las carnes mórbidas de su brazo y tira de ella hasta **ponerla a caminar**. (Sánchez-Andrade, Cristina, *Bueyes y rosas dormían*, 2001).

‘He lifts her from the seat, hooks her hands between the morbid flesh of his arm and pulls her up until **he makes her walk**’.

In (17a), the causer (*Alcina*) induces the causee (*los hombres* ‘men’) to get involved in another event (*gozar de sus regalos* ‘enjoy her gifts’) by verbal and psychological means, namely *por engaño* (‘deceitfully’). As the causee performs the caused event autonomously, likely at a different time than the causing event, a situation like the one in (17a) can be categorized as indirect causation. On the contrary, in (17b), the causer exerts a physical force on the causee (*la* ‘her’), who, despite being animate, lacks autonomy and control over the caused event (*caminar* ‘walk’), which occurs (almost) simultaneously with the causing event. Therefore, examples like (17b) should be categorized as direct causation. A qualitative analysis of our corpus reveals that, when causative micro-constructions with *llevar* and *traer* feature two animate participants, causation is always achieved by verbal, social or psychological means, as in (17a), while situations of physical causation such as the one in (17b) are not documented. Therefore, all examples in our corpus with two animated participants are considered cases of indirect causation.

Another possibility is that both the causer and the causee are inanimate entities, as in (18). In these cases, the causer exerts a physical force on the causee. The situation results in the caused event, without the causer nor the causee having any control over it. Since the causee does not act autonomously, and the caused event is simultaneous to the causing event, the situation described is closer to direct causation and is categorized as such.

(18) *El mismo mecanismo de rotación o enrollamiento lleva el fórnix gástrico* a colocarse en contacto con el tercio inferior del esófago. (M. Díaz Rubio, *Lecciones de patología y clínica médica. Aparato digestivo*, 1964).

‘The same mechanism of rotation or coiling **leads** the gastric fornix to stand into contact with the lower third of the esophagus’.

It follows from this that the animacy of the participants is an important parameter to characterize the type of causation expressed by a construction. As mentioned before (supra Section 1), the causative micro-construction with *llevar* has been placed within the scope of indirect causation (Da Silva 2004, 2012). Still, it has been noted that the emergence of a new grammatical construction usually coincides with an expansion of the construction into new,

less prototypical semantic types, as a consequence of the abstraction of the schema (Barðdal 2008; Traugott and Trousdale 2013). If we assume, following Da Silva (2004, 2012), that the micro-constructions with *llevar* and *traer* express prototypically indirect causation, it is reasonable to think that, in Old Spanish, they were already strongly associated with events of indirect causation, and that over time they expressed more diverse types of causation and more flexibly admitted different types of participants.

With regard to the nature of the causer, if the formation of the micro-constructions under study has led to the development of a new causative subschema, it is expected that they initially selected inanimate and later more easily accepted animate causers (NP1). Table 6 presents the results of this classification.¹⁵

Table 6. Animacy of the causer (NP1).

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----------------------------|---------|----------|----------|----------|----------|----------|-----------|-----------|
| <i>Llevar</i> NP1 [anim] | - | 1 (100%) | 1 (11%) | 5 (12%) | 8 (17%) | 2 (12%) | 11 (4%) | 27 (3%) |
| <i>Llevar</i> NP1 [inanim] | - | - | 8 (89%) | 38 (88%) | 40 (83%) | 14 (88%) | 242 (96%) | 748 (97%) |
| Total <i>llevar</i> | - | 1 | 9 | 43 | 48 | 16 | 253 | 775 |
| <i>Traer</i> NP1 [anim] | 3 (60%) | 5 (38%) | 20 (39%) | 34 (43%) | 10 (27%) | - | 1 (8%) | - |
| <i>Traer</i> NP1 [inanim] | 2 (40%) | 8 (62%) | 31 (61%) | 45 (57%) | 27 (73%) | 2 (100%) | 12 (92%) | 1 (100%) |
| Total <i>traer</i> | 5 | 13 | 51 | 79 | 37 | 2 | 13 | 1 |

The results show that both micro-constructions select mostly inanimate causers, although not entirely to the same degree. The micro-construction with *traer* appears with animate causers with some frequency until the 17th century. For instance, in the 16th century, cases with animate causers account for 43% of the total number of cases. However, from the 17th century on, the number of animate causers decreases, coinciding with the decline in productivity of the micro-construction. With respect to the micro-construction with *llevar*, the proportion of animate causers is much lower, ranging between 11% and 17% between the 15th and 18th centuries and decreasing drastically in the 19th and 20th centuries. These data suggest that, rather than an expansion of the construction to new semantic domains, there was a specialization of the micro-constructions under study towards prototypical indirect causation, since they select inanimate causers with increasing rigidity.

Since prototypical indirect causation involves an animate causee (Engels and Comer 2020), it is expected that the causative micro-constructions with *llevar* and *traer* initially selected animate causees and that, over time, there was an increase in inanimate ones. The semantic classification of the causee (NP2) yields the results presented in Table 7.

Table 7. Animacy of the causee (NP2).

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----------------------------|----------|----------|----------|----------|----------|----------|-----------|-----------|
| <i>Llevar</i> NP2 [anim] | - | 1 (100%) | 6 (67%) | 43 (90%) | 44 (90%) | 16 (84%) | 245 (96%) | 790 (98%) |
| <i>Llevar</i> NP2 [inanim] | 1 (100%) | - | 3 (33%) | 5 (10%) | 5 (10%) | 3 (16%) | 9 (4%) | 17 (2%) |
| Total <i>llevar</i> | 1 | 1 | 9 | 48 | 49 | 19 | 254 | 807 |
| <i>Traer</i> NP2 [anim] | 5 (83%) | 11 (85%) | 56 (95%) | 82 (95%) | 35 (95%) | 2 (100%) | 14 (100%) | 1 (100%) |
| <i>Traer</i> NP2 [inanim] | 1 (17%) | 2 (15%) | 3 (5%) | 4 (5%) | 2 (5%) | - | - | - |
| Total <i>traer</i> | 6 | 13 | 59 | 86 | 37 | 2 | 14 | 1 |

The results show that, from the beginning, both micro-constructions were mostly formed with animate causees. Moreover, in both cases, there is a clear trend towards an increasingly rigid selection of animate causees. In concrete, in the micro-construction with *traer*, the proportion of inanimate causees is reduced from 17% (13th century) to 0% (18th to 20th centuries), while in the case of *llevar*, it is reduced from 33% (15th century) to 2% (20th century). These data reinforce the idea that, rather than expanding their scope, the

micro-constructions under study became increasingly specialized in order to express events of indirect causation.

Finally, to get a clearer picture of the semantic profile of the micro-constructions with *llevar* and *traer* as a whole, the examples are classified according to the animacy of both participants (NP1 and NP2) in Table 8.

Table 8. Animacy of the causer (NP1) and the causee (NP2).

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---------|----------|----------|----------|----------|----------|-----------|-----------|
| <i>Llevar</i> NP1 [anim] + NP2 [inanim] | - | - | 1 (11%) | - | 2 (4%) | - | 6 (2%) | 7 (1%) |
| <i>Llevar</i> NP1 [inanim] + NP2 [inanim] | - | - | 2 (22%) | 4 (9%) | 3 (6%) | 1 (6%) | 3 (1%) | 10 (1%) |
| <i>Llevar</i> NP1 [anim] + NP2 [anim] | - | 1 (100%) | - | 5 (12%) | 6 (13%) | 2 (13%) | 5 (2%) | 20 (3%) |
| <i>Llevar</i> NP1 [inanim] + NP2 [anim] | - | - | 6 (67%) | 34 (79%) | 37 (77%) | 13 (81%) | 239 (94%) | 758 (95%) |
| Total <i>llevar</i> | - | 1 | 9 | 43 | 48 | 16 | 253 | 795 |
| <i>Traer</i> NP1 [anim] + NP2 [inanim] | - | - | 1 (2%) | 1 (1%) | 1 (3%) | - | - | - |
| <i>Traer</i> NP1 [inanim] + NP2 [inanim] | - | 2 (15%) | 2 (4%) | 1 (1%) | 1 (3%) | - | - | - |
| <i>Traer</i> NP1 [anim] + NP2 [anim] | 3 (60%) | 5 (38%) | 19 (37%) | 33 (42%) | 9 (24%) | - | 1 (8%) | - |
| <i>Traer</i> NP1 [inanim] + NP2 [anim] | 2 (40%) | 6 (46%) | 29 (57%) | 44 (56%) | 26 (70%) | 2 (100%) | 12 (92%) | 1 (100%) |
| Total <i>traer</i> | 5 | 13 | 51 | 79 | 37 | 2 | 13 | 1 |

The results presented in Table 8 confirm the observed tendency towards semantic specialization. Up until the 16th century, the micro-construction with *traer* is easily employed to express situations with an inanimate causer and an animate causee, as in (19a), and also with two animate participants (19b). The frequency of examples with both animate participants ranges from 60% to 37% until the 16th century. From the 17th century on, there is a drop with no return.

(19) a. La natura spirital traer al omne a saber todas las cosas segunt son. (Libro de los buenos proverbios que dijeron los filósofos y sabios antiguos, 1250).

‘The spiritual nature **leads** men to know all things as they are’.

b. Traidor, que con tu ley halaguera me engañaste, y *has traído* a dexar la verdadera. (Jorge Manrique, *Los fuegos*, 1469–1479).

‘Traitor, who with your flattering law have deceived me, and have **brought** me to forsake the true one’.

In contrast, the micro-construction with *llevar* with two animate participants (20) was always rare, with a proportion ranging between 12% and 13% between the 16th and 18th centuries. In the 19th and 20th centuries, the frequency of this type declines further to 2% and 3%.

(20) Algun dulce el maestro al niño dando, lo *lleva así á estudiar* más fácilmente. (Manuel María Arjona, *Poesías*, 1790–1820).

‘The teacher gives some candy to the child, so he **leads** him to study more easily.’

With both micro-constructions, the possibility of expressing causative events with both inanimate participants, as shown supra in (18), or with an animate causer and an inanimate causee, as shown supra in (16a), which are types of events related to direct causation, was always infrequent. However, the proportion of such cases drops even further in the 19th and 20th centuries (0–2%).

These data show that the causative micro-constructions with *llevar* and *traer* specialized in events of indirect causation from the earliest documentations on. In this, the micro-constructions with *llevar* and *traer* differ from others, such as the variants with *hacer* and *poner*, which more readily allow the expression of both direct and indirect causation (Engiels and Comer 2020). What is more, far from diversifying their semantic profile, the micro-constructions with *llevar* and *traer* increasingly specialized towards the prototype of indirect causation, which involves an inanimate causer and an animate causee. This type of examples raised from 40% to 100% for *traer* and from 67% to 95% for *llevar*. This suggests that a subschema with CAM verbs developed within the causative construction

with a specialized function, namely the expression of indirect causation. Since indirect causation is associated with less grammaticalized structures (Shibatani and Pardeshi 2002), the further specialization of the subschema in indirect causation events explains why there has been no increase over time of its degree of unithood (see Section 4.2).

However, the ease with which the *traer* micro-construction accepted animated causers up to the 16th century represents an important difference respect to the micro-construction with *llevar*, which from the beginning was used in a more restricted way for prototypical indirect causation. This points to the fact that the semantic profile of the micro-construction with *traer* was progressively attracted to that of the micro-construction with *llevar* in Classical Spanish, a period in which *llevar* started to lead the subschema. The leadership of the micro-construction with *llevar* is explained by the semantic changes experienced by *llevar* and *traer* as motion verbs, which made the meaning of the micro-construction with *llevar* more transparent, and ultimately caused the decline in the micro-construction with *traer*.

4.3.2. Semantics of the Caused Event

If speakers abstracted the meaning of the causative micro-constructions with *llevar* and *traer* and created a new subschema, it could be expected that, over time, it would extend to new semantic contexts (Barðdal 2008; Traugott and Trousdale 2013). Moreover, the removal of the original semantic constraints is associated with the grammaticalization of the construction, as a consequence of the loss of its semantic integrity (Lehmann 2015).

In order to approach the lexical-semantic fields with which the micro-constructions under study are associated, the infinitives in the dataset are labelled on the basis of the ADESSÉ classification (García-Miguel et al. 2010). This classification distinguishes six verbal macro-categories that are further subdivided into more specific categories, up to three levels of analysis. For example, a macro-class of mental verbs is distinguished, which at a second level of analysis is divided into verbs of feeling (e.g., *gustar* ‘to like’), verbs of perception (e.g., *ver* ‘to see’) and verbs of cognition (e.g., *pensar* ‘to think’). On their turn, some verbs of cognition are ascribed, at a third level of analysis, to the more specific classes of verbs of knowledge (e.g., *saber* ‘to know’) and verbs of belief (e.g., *creer* ‘to believe’). In order to achieve a precise level of description, the most detailed level of analysis is adopted. Table 9 shows the number of different semantic classes presented by the infinitives in each century.

Table 9. Number of different semantic classes of the infinitive per century.

| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------------|----|----|----|----|----|----|----|----|
| <i>Llevar</i> | 1 | 1 | 6 | 20 | 21 | 12 | 39 | 49 |
| <i>Traer</i> | 5 | 8 | 21 | 24 | 15 | 2 | 11 | 1 |

These data confirm that, after their emergence, the causative micro-constructions with *llevar* and *traer* became associated with infinitives of increasingly varied meanings. In the case of *llevar*, this trend extends from the 13th century to the 20th century, except for a decline in the 18th century, which may be due to the smaller number of examples. On the other hand, with respect to the micro-construction with *traer*, from the 17th century onwards, the semantic diversity of infinitives is reduced, coinciding with the general decline in the productivity of the micro-construction. The progressive diversification of the semantic classes of the infinitives points to the fact that a constructional schema is developing over time.

In addition, it is worth considering (a) which are the semantic classes expressed by the infinitives, (b) whether the semantic classes have changed over time and (c) whether they are different in the two micro-constructions under study. As was mentioned before, the micro-constructions derive from expressions that typically indicate a metaphorical

movement towards a state conceptualized as an abstract place, as in example (7a), which is now recovered as (21) (see Section 2).

(21) El uno de tus amigos es aquel que te tuelle del mal y te **lieva al bien** (*Libro de los buenos proverbios que dijeron los filósofos y sabios antiguos*, 1250)

‘Your friend is the one who takes you out of evil and **leads** you to good’. Therefore, they are expected to be initially associated with attributive and other stative events, and to later include other types of events and actions. Table 10 shows the most frequent semantic classes of the infinitive per century and per micro-construction.¹⁶

Table 10. Most frequent semantic classes of the infinitive per century.

| | <i>llevar</i> | | <i>traer</i> | |
|----|---------------|-----|---------------|----|
| | Class INF | # | Class INF | # |
| 13 | | | knowledge | 2 |
| 14 | | | attribution | 3 |
| | | | light verbs | 3 |
| | | | knowledge | 2 |
| 15 | knowledge | 3 | light verbs | 9 |
| | displacement | 2 | communication | 5 |
| | | | life | 5 |
| | | | attribution | 4 |
| 16 | | | permission | 4 |
| | perception | 10 | attribution | 14 |
| | feeling | 5 | light verbs | 12 |
| | attribution | 4 | feeling | 8 |
| | communication | 4 | life | 8 |
| 17 | knowledge | 3 | competition | 5 |
| | perception | 12 | attribution | 6 |
| | communication | 4 | perception | 5 |
| | life | 4 | feeling | 5 |
| | attribution | 3 | life | 5 |
| 18 | possession | 3 | activity | 3 |
| | attribution | 3 | | |
| | communication | 3 | | |
| | perception | 3 | | |
| 19 | feeling | 2 | | |
| | perception | 25 | attribution | 2 |
| | communication | 19 | communication | 2 |
| 20 | | | light verbs | 2 |
| | cognition | 16 | | |
| | belief | 13 | | |
| 20 | communication | 110 | | |
| | perception | 73 | | |
| | cognition | 55 | | |
| | belief | 51 | | |
| | knowledge | 49 | | |

Table 10 reveals important differences between the micro-constructions and between different language stages. From its first attestations, the micro-construction with *llevar* is frequently associated with mental verbs, especially with the subclasses of perception verbs, such as *buscar* (‘search’) (22a), and verbs of cognition, such as *pensar* (‘think’) (22b), which is the most frequent infinitive in the 19th and 20th centuries. Moreover, from the 16th century

on, and especially in the 20th century, it is often employed for communication events, as with the verb *expresar* ('express') (20c).

(22) a. Hay en mi corazón una atracción secreta, que *me lleva* buscaros. (Emilio Castelar, *Ernesto: novela original de costumbres*, 1855).

'There is in my heart a secret attraction, which **leads** me to look for you'.

b. Estas cosas *le llevan* a uno a pensar en la vida. (Miguel Delibes, *Diario de un emigrante*, 1958).

'These things **lead** one to think about life'.

c. El sentimiento religioso *lleva al hombre a expresar* su retorno hacia Dios. (Rafael Alcocer, *Iniciación litúrgica: la misa*, 1935).

'Religious sentiment **leads** men to express their return to God'.

Regarding the micro-construction with *traer*, unlike *llevar*, it is very often associated with light verbs, particularly with *hacer* ('make') in examples such as (21), in all epochs except the 17th century.

(23) No diga el que ahorcan que su hado lo traxo a morir aquella muerte, [. . .], que lo que *los trae a hazer tan ruyñ fin* de su vida es su poca consideración. (Antonio de Torquemada, *Jardín de flores curiosas*, 1569).

'Do not say that the one who is hanged that his fate brought him to die that death, [. . .], that what **leads** them to make such a dastardly end to their life is their lack of consideration'.

Moreover, in their early stages, both micro-constructions frequently expressed events of attribution, particularly with the verb *ser* 'to be' (24). However, while in the case of *traer* this association is much more frequent and remains constant throughout history (14th to 19th centuries), in the case of *llevar*, it weakened after the 18th century.

(24) a. Que envidia nin cobdicia de plata nin dineros non les busquen nin *trayan a ser* fallesçederos. (Pero López de Ayala, *Rimado de Palacio*, 1378–1406).

'That envy nor greed for silver or money do not seek them nor **lead** them to be perishable'.

b. No he tenido yo la culpa, sino quien *le ha llevado de la mano a ser* tan loco. (Lope de Vega, *La prudente venganza*, 1623).

'It was not my fault, but of the one who **led** him by the hand to be so crazy'.

This seems to indicate that both micro-constructions were at some point linked to attributive caused events, as it was expected because of their origin, although to different extents. Over time, both expanded to different semantic domains, but the variant with *traer* remained more closely linked to stative event types. This suggests a lower degree of grammaticalization of the micro-construction with *traer*, which does not experience a loss of its semantic integrity as markedly as *llevar*.

5. Discussion and Conclusions

This paper has studied the diachrony of causative micro-constructions with the verbs *llevar* ('take') and *traer* ('bring'). The analysis of seven grammaticalization parameters has shown that the two micro-constructions share a series of formal and semantic features that differentiate them from other causative micro-constructions. As an outcome of the analysis, we propose that a specific subschema including verbs of Caused Accompanied Motion (CAM) has developed as part of the overarching causative schema.

The subschema is formally characterized by the presence of the preposition *a* before the infinitive, a feature that has become completely fixed since the 18th century, and by a low degree of unithood, implying a low degree of syntactic incorporation between the causative verb and the infinitive. In concrete, it has been shown that the postposition of the causee to the infinitive is very infrequent in the subschema with CAM verbs, as opposed to

what was observed for other causatives with *hacer* ('to make') and *poner* ('to put') (Enghels and Comer 2020, p. 176). It follows, therefore, that, compared to related variants, the causative subschema with CAM verbs has a relatively low degree of incorporation.

From a semantic point of view, it has been shown that the subschema with CAM verbs prototypically expressed indirect causation. Moreover, it increasingly specialized in prototypical indirect causation events, involving an inanimate causer and an animate causee. Again, this contrasts with the subschema with *putting* verbs, which specialized in the expression of direct causation, and with the micro-construction with *hacer*, which presents greater semantic versatility (Enghels and Comer 2020).

The two micro-constructions under study also display some diverging characteristics. The micro-construction with *llevar* more rigidly selects inanimate causers than the micro-construction with *traer*, which also admits animate causers with some ease. Moreover, the micro-construction with *llevar* is frequently associated with infinitives expressing perception and communication events, while the micro-construction with *traer* is more strongly associated with attributive events, such as those expressed by *ser* ('to be'). However, most of the grammaticalization parameters analyzed point to a similar degree of grammaticalization of the micro-constructions, which has remained relatively stable over time.

From the perspective of their productivity, the subschema with CAM verbs developed following a S-shaped curve, with a slow beginning followed by a rapid expansion (Nevalainen 2015). However, it has been shown that the micro-construction with *traer* dominated the subschema until the 15th century, while the variant with *llevar* developed later and more slowly. This has been explained by the Aktionsart features of *llevar* as motion verb in Old Spanish, which hindered its incorporation into the causative construction. In contrast, from the 16th century onwards, the consolidation of the deictic meaning of *traer* as motion verb caused the micro-construction with *traer* to become less semantically transparent, and consequently, its productivity decreased. On the contrary, the loss of the original aspectual features of *llevar* made the micro-construction with *llevar* more transparent, so in Classical Spanish, it started leading the subschema. This also explains why after the 18th century the micro-construction with *llevar* expanded to new semantic domains (e.g., caused events of communication), while the micro-construction with *traer* became residual, especially in the 20th century.

The above-mentioned changes illustrate that the semantic changes experienced by *llevar* and *traer* as full verbs determined the productivity of the causative micro-constructions under study. These findings suggest that the full meanings of verbs and their grammaticalized uses are strongly connected in the minds of speakers. Therefore, as Garachana Camarero and Hernández Díaz (2020) already pointed out in their study on verbal periphrases in Spanish, we argue that studies on grammaticalization should pay enough attention to possible diachronic changes in the full lexical meaning of verbs, since they may impact the evolution of grammaticalized constructions.

To conclude, in future research, it would be interesting to study whether the observed tendencies also apply to other causative micro-constructions with CAM verbs, such as *aducir* ('carry') (25a) and *conducir* ('drive, lead') (25b).

(25) a. Por que el entendimiento los **aduxiesse** a connosçer las cosas ssegunt que eran primeramente a Dios (Alfonso X, *Setenario*, 1252–1270).

'That the understanding **lead** them to know things as they were firstly to God'.

b. Esta noticia naturalmente *me conduce* a *rectificar* otra. (Benito Jerónimo Feijoo, *Teatro Crítico Universal*, 1734).

'This news naturally **leads** me to rectify another'.

Author Contributions: Conceptualization, J.T.S. and R.E.; Methodology, J.T.S. and R.E.; Validation, J.T.S. and R.E.; Formal analysis, J.T.S.; Investigation, J.T.S. and R.E.; Data curation, J.T.S.; Writing—original draft, J.T.S. and R.E.; Writing—review & editing, J.T.S. and R.E.; Visualization, J.T.S.; Supervision, R.E.; Funding acquisition, J.T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Ministry of Universities grant for university teacher training number FPU19/01505.

Data Availability Statement: The research data of this study is available in TROLLing (Torres Soler and Enghels 2023). Torres Soler, Julio, and Renata Enghels. 2023. Replication Data for: From Motion to Causation: The Diachrony of the Spanish Causative Constructions with *traer* ('Bring') and *llevar* ('Take'). DataverseNO. <https://doi.org/10.18710/BQ0PDU>.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ The parenthesis in the slot of the preposition indicates that it may appear or not.
- ² In this paper we do not adopt the term *verbal periphrasis*, which many authors use in a restricted sense excluding the causative construction. Instead, we opt for the more comprehensive terms of *construction* and *micro-construction* as understood by CxG. It is assumed, in any case, that we are dealing with verbal expressions that present a certain degree of grammaticalization.
- ³ This is the most frequent and oldest meaning of the verbs (Torres Soler 2020).
- ⁴ The following lines provide a summary of the main ideas exposed in Torres Soler (2023) and relevant to explain the diachrony of the causative micro-constructions under study.
- ⁵ An anonymous reviewer correctly warns that this example comes from a document with low ecdotic reliability. Even so, it illustrates in a very clear way the conceptual relationship between the causative construction and the metaphorical expressions from which it derives, and this is why we decided to include the example.
- ⁶ See <https://www.rae.es/banco-de-datos/cdh> (accessed on 15 March 2022).
- ⁷ This implies that examples with more than three elements interpolated between the causative verb and the infinitive are not included in our corpus. However, it does not affect our analysis, which aims to clarify whether it is possible or not to interpolate words between the causative verb and the infinitive. The number of interpolated elements is not relevant for the purposes of this work.
- ⁸ As other diachronic corpora of Spanish, the CDH contains a much lower number of texts from the 18th century than from earlier and later centuries. Specifically, in the geographic area of Spain it contains 12,372,992 words in the 18th century, whereas the amount of words is 35,471,967 in the 17th century and 39,543,425 in the 19th century. This quantitative data disparity may cause some unexpected results when observing diachronic trends, especially when the number of tokens is low. This will be taken into account during the interpretation of the results.
- ⁹ Several arguments have been made in favour of considering *coming* verbs as the marked elements of the pair of deictic motion verbs (Ricca 1993). One of them is the fact that the *going* verb is the one used in deictically neutral expressions or expressions in which the Goal is interpreted in a generic way, as in the expression *me gusta ir al cine* ('I like to go to the cinema'), in which the verb *ir* ('go') can be used even if the speaker is in the cinema at the moment of speech. Another argument is that, sometimes, the use of the *coming* verb generates more rigid inferences about the presence/absence of the speaker at the Goal than the same utterance with the *going* verb. An example is *¿vendrás a la fiesta esta noche?* ('Will you come to the party tonight?'), from which it is inferred that the speaker is at the place of the party at the moment of speech or will be there at the moment of reference, as opposed to the expression *¿Irás a la fiesta esta noche?* ('Will you go to the party tonight?'), from which neither the presence nor the absence of the speaker at the party can be inferred.
- ¹⁰ A similar variation in the prepositional slot has been documented in the causative micro-constructions with *poner* and *meter* (Enghels and Comer 2020).
- ¹¹ Due to the low density of causative constructions in the highly extensive list of search results, we retained only cases in which the causative verb and the preposition *en* or the causative verb and the infinitive were contiguous. Consequently, the frequencies of the prepositional variants with *en* and \emptyset are not completely comparable to those with the preposition *a*, although they do provide a more concrete idea of their diachronic development.
- ¹² It goes without saying that cases in which the causee is expressed through a clitic are excluded (e.g., *El miedo les llevó a huir* 'fear made them run away').
- ¹³ Cases without any adjunct susceptible of being intercalated are excluded from this subsample.
- ¹⁴ The different categorial nature of the causee (NP2) and the adjuncts entails a different degree of syntactic mobility. As a consequence, the comparison of the degree of interpolation of both types of elements with each other does not yield relevant results. Similarly, Comer (2020, p. 355) states that "the grammatical class (nuclear or non-nuclear) of the interpolated constituents

and the degree of grammaticalization do not correlate”. Instead, taking the interpolation of different grammatical classes as independent variables allows to test whether the degree of lexical interpolation between the causative verb and the infinitive is subject to diachronic variation or variation across micro-constructions.

- 15 Human participants, living animals, institutions (e.g., the Ministry of Universities), and personified divine or mythological beings (e.g., Satan) have been classified as animate. Abstract concepts, objects and nominal expressions designating actions (e.g., a thief’s escape) are categorized as inanimate participants. Passive sentences in which there is no explicit causer are excluded (e.g., *Así los niños son llevados a obedecer* ‘In this way children are led to obey’).
- 16 We included the 5 most frequent semantic classes in each century, and only those that appear more than once.

References

- Alfonso Vega, Milagros. 1997. *Construcciones causativas de infinitivo en el español medieval: Estructura y evolución*. Ph.D. thesis, El colegio de México, Ciudad de México, Mexico.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins Publishing Company.
- Cano-Aguilar, Rafael. 1977. Las construcciones causativas en español (continuación). *Boletín de la Real Academia Española* 57: 323–35.
- Cano-Aguilar, Rafael. 1988. *El español a través de los tiempos*. Madrid: Arco/Libros.
- Cifuentes Honrubia, José Luis. 2008. Verbos de remoción y alternancias locales en español. *ELUA* 22: 37–64. [CrossRef]
- Comer, Marie. 2020. *Desplazamiento, incoación y causación: Sobre la cuasi-sinonimia y gramaticalización de «poner» y «meter» en español*. Berlin: De Gruyter Mouton.
- Correia Saavedra, David. 2021. *Measurements of Grammaticalization: Developing a Quantitative Index for the Study of Grammatical Change*. Berlin: De Gruyter Mouton.
- Da Silva, Augusto Soares. 2004. Verbos y construcciones causativas analíticas en portugués y español. *ELUA* 2: 581–98. [CrossRef]
- Da Silva, Augusto Soares. 2012. Stages of grammaticalization of causative verbs and constructions in Portuguese, Spanish, French and Italian. *Folia Linguistica* 46: 513–52. [CrossRef]
- Davies, Mark E. 1995. The Evolution of the Spanish Causative Construction. *Hispanic Review* 63: 57–77. [CrossRef]
- Engels, Renata, and Eugene Roegiest. 2014. Contrasting the syntax and semantics of negative causation: The apparent similarity of Spanish and Portuguese. *Languages in Contrast* 14: 278–305. [CrossRef]
- Engels, Renata, and Marie Comer. 2018. Evaluating grammaticalization and constructional accounts: The development of the inchoative construction with put verbs in Spanish. In *Grammaticalization Meets Construction Grammar*. Edited by Evie Coussé, Joel Olofsson and Peter Andersson. Amsterdam: John Benjamins Publishing Company, pp. 107–33.
- Engels, Renata, and Marie Comer. 2020. La red construccional de perifrasis causativas: Definición, comparación sincrónica y evolución diacrónica. In *La evolución de las perifrasis verbales en español: Una aproximación desde la gramática de construcciones diacrónica*. Edited by Garachana Mar. Bern: Peter Lang.
- Garachana Camarero, Mar, and Axel Hernández Díaz. 2020. From semantics to grammar: Lexical substitution in the evolution of verbal periphrases haber/tener + infinitive. In *Changes in Meaning and Function: Studies in Historical Linguistics with a Focus on Spanish*. Edited by Jorge Fernández Jaén and Herminia Provencio Garrigós. Amsterdam: John Benjamins Publishing Company, pp. 77–108.
- García-Miguel, José M., Gael Vaamonde, and Fita González Domínguez. 2010. ADESSE, a Database with Syntactic and Semantic Annotation of a Corpus of Spanish. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*. Edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner and Daniel Tapias. Valetta: European Language Resources Association (ELRA), pp. 1903–10.
- Guilquin, Gaëtanelle. 2010. *Corpus, Cognition and Causative Constructions*. Amsterdam: John Benjamins Publishing Company.
- Hamplová, Sylva. 1970. *Algunos problemas de la voz perifrástica pasiva y las perifrasis factitivas en español*. Prague: Instituto de Lengua y Literaturas de la Academia Checoslovaca de Ciencias.
- Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization*. Oxford: Oxford University Press.
- Hopper, Paul. 1991. On some principles of grammaticization. In *Approaches to Grammaticalization: Volume I. Theoretical and Methodological Issues*. Edited by Elizabeth Closs Traugott and Bernd Heine. Amsterdam: John Benjamins Publishing Company, pp. 17–36.
- Kemmer, Suzanne, and Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5: 115–56. [CrossRef]
- Kopecka, Anetta, and Marine Vuillermet. 2021. Source-Goal (a)symmetries across languages. *Studies in Language* 45: 2–35. [CrossRef]
- Lakoff, George, and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Lamiroy, Béatrice. 1994. Les syntagmes nominaux et la question de l’auxiliarité. *Langages* 115: 64–75. [CrossRef]
- Lavale, Ruth. 2013. *Verbos denominales causativos en español actual*. Ph.D. thesis, University of Alicante, San Vicente del Raspeig, Spain.
- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*, 3rd ed. Berlin: Language Science Press.
- Maldonado, Ricardo. 2007. Soft causatives in Spanish. In *On Interpreting Construction Schemas: From Action and Motion to Transitivity and Causality*. Edited by Nicole Delbecque and Bert Cornillie. Berlin: De Gruyter Mouton, pp. 229–60.

- Margetts, Anna, Katharina Haude, Nikolaus P. Himmelmann, Dagmar Jung, Sonja Riesberg, Stefan Schnell, Frank Seifart, Harriet Sheppard, and Claudia Wegener. 2022. Cross-linguistic patterns in the lexicalisation of bring and take. *Studies in Language* 46: 934–93. [CrossRef]
- Nevalainen, Terttu. 2015. Descriptive adequacy of the S-curve model in diachronic studies of language change. *Studies in Variation, Contacts and Change in English*. 16. Available online: <https://varieng.helsinki.fi/series/volumes/16/nevalainen/#references> (accessed on 15 March 2023).
- Octavio de Toledo Huerta, Álvaro. 2016. Aprovechamiento del CORDE para el estudio sintáctico del primer español moderno (ca. 1675–825). In *Lingüística de corpus y lingüística histórica iberorrománica*. Edited by Johannes Kabatek. Berlin: De Gruyter, pp. 57–89.
- RAE. 2009. *Nueva gramática de la lengua española*. Buenos Aires: Espasa Calpe.
- Ricca, Davide. 1993. *I verbi deittici di movimento in Europa: Una ricerca interlinguistica*. Firenze: La Nuova Italia.
- Rodríguez Molina, Javier. 2010. La gramaticalización de los tiempos compuestos en español antiguo: Cinco cambios diacrónicos. Ph.D. thesis, Universidad Autónoma de Madrid, Madrid, Spain.
- Roegiest, Eugeen, and Renata Enghels. 2008. La reducción oracional en la construcción factitiva española. In *Lenguas en diálogo: El iberorromance y su diversidad lingüística y literaria: Ensayos en homenaje a Georg Bossong*. Edited by Hans-Jörg Döhla and Raquel Montero-Muñoz. Madrid: Iberoamericana, pp. 289–312.
- Shibatani, Masayoshi, and Prashant Pardeshi. 2002. The causative continuum. In *The Grammar of Causation and Interpersonal Manipulation*. Edited by Masayoshi Shibatani. Amsterdam: John Benjamins Publishing Company, pp. 85–126.
- Stefanowitsch, Anatol, and Ada Rohde. 2004. The goal bias in the encoding of motion events. In *Studies in Linguistic Motivation*. Edited by Günter Radden and Klaus-Uwe Panther. Berlin: Mouton de Gruyter, pp. 249–67.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics: Typology and Process in Concept Structuring*. Cambridge: The MIT Press, vol. 2.
- Torres Cacouillos, Rena. 2000. *Grammaticization, Synchronic Variation, and Language Contact. A Study of Spanish Progressive -ndo Constructions*. Amsterdam: John Benjamins Publishing Company.
- Torres Soler, Julio. 2020. La polisemia del verbo “llevar” a la luz de la semántica diacrónica de prototipos. *Tonos digital* 39: 1–28.
- Torres Soler, Julio. 2023. *How to Become a Deictic Motion Verb? The Case of Spanish llevar ('Take') and traer ('Bring')*. San Vicente del Raspeig: University of Alicante, Manuscript in preparation.
- Traugott, Elisabeth, and Bernd Heine. 1991. Introduction. In *Approaches to Grammaticalization: Volume I. Theoretical and Methodological Issues*. Edited by Elisabeth Closs Traugott and Bernd Heine. Amsterdam: John Benjamins Publishing Company, pp. 1–16.
- Traugott, Elisabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Treviño, Esthela. 1994. *Las causativas del español con complemento infinitivo*. Ciudad de México: El colegio de México.
- Vivanco, Margot. 2019. Haciendo encajar las piezas en las causativas analíticas con ‘hacer’. *Verba: Anuario Galego De Filoloxía* 46: 69–108. [CrossRef]
- Vivanco, Margot. 2020. ¿Qué hay entre el control y la reestructuración? Sobre la construcción <poner algo ~ a alguien a + infinitivo>. *Revista española de lingüística* 50: 13–242.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Constructing Meaning: Historical Changes in MIHI EST and HABEO Constructions in Romanian

Mihaela Ilioiaia

Faculty of Arts and Philosophy, Department of Linguistics, Ghent University, Campus Boekentoren, Blandijnberg 2, 9000 Ghent, Belgium; mihaela.ilioaia@ugent.be

Abstract: In this article, I address the evolution of the competition between two Latin patterns, *habeo* and *mihi est*, in Romanian. As opposed to the other Romance languages, which replace the *mihi est* pattern with *habeo* in possessor and experiencer contexts, Romanian maintains both Latin patterns. The general evolution of these patterns in the Romance languages is well known, however, a detailed usage-based account is currently lacking. Building on the theoretical findings on the role of functional competition in linguistic change, the rivalry between the two patterns in Romanian has already been argued to have settled in terms of differentiation, with each of the two forms specializing in different functional domains by Vangaever and Ilioiaia in 2021 in their study “*Specialisation through competition: habeo vs. mihi est from Latin to Romanian*”. With this idea as a starting point, I investigate, by means of a diachronic corpus study, whether the dynamics in the inventory of state nouns occurring in these constructions can affect their evolution and productivity. The preliminary results show that this is indeed the case. Concomitantly, I explore whether the historical changes that the two patterns have undergone over the centuries can be described in terms of grammaticalization, constructionalization, or in terms of constructional change.

Keywords: MIHI EST; HABEO; competition; constructionalization; constructional change; Romanian; Latin; substitution; differentiation; functional competition

Citation: Ilioiaia, Mihaela. 2024. Constructing Meaning: Historical Changes in MIHI EST and HABEO Constructions in Romanian. *Languages* 9: 38. <https://doi.org/10.3390/languages9020038>

Academic Editor: Jeanine Treffers-Daller

Received: 30 October 2023
Revised: 10 January 2024
Accepted: 11 January 2024
Published: 24 January 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper deals with the two Romanian constructions illustrated in (1) and (2) below. The construction in (1) is an instantiation of a MIHI EST construction (cf. Ilioiaia 2021; Ilioiaia and Van Peteghem 2021), in which the verb *fi* ‘be’ occurs with a noun denoting a state (henceforth ‘state noun’) and a dative experiencer, a combination that conveys to the construction the meaning of a psychological or a physiological state. The construction in (2) illustrates the HABEO construction and expresses exactly the same meaning, this time from the combination of the verb *avea* ‘have’ with a state noun and an experiencer in the nominative.

(1) *Băiatului îi e frică de poliție*
boy him.DAT is fear of police
‘The boy is afraid of police.’

(2) *Io tot am frică la dentist*
I still have fear.ACC=NOM at dentist
‘I am still afraid when I go to the dentist.’ (ioanaspune.ro, accessed on 5 May 2018)

In Romanian, the MIHI EST, as well as the HABEO constructions, developed from the inherited Latin patterns (cf. Benveniste 1966, p. 197; Bauer 1996). They have both been preserved in Romanian, while, in other Romance languages, the HABEO construction entirely replaces the MIHI EST construction. The two inherited constructions are in competition in Romanian (Ilioiaia 2021; Vangaever and Ilioiaia 2021; Ilioiaia 2023), and they are argued to have already been in competition with each other in Latin (cf. Fedriani 2011; Baldi and Nuti 2010, pp. 260–61). With respect to the historical changes that these two constructions

underwent, I will explore whether they can be accounted for in terms of grammaticalization, constructionalization, or in terms of constructional change.

The general evolution of these patterns is well known; however, a detailed account is currently lacking. This may be due to the absence of a coherent theoretical framework to properly assess this evolution and the lack of quantitatively oriented corpus-based studies addressing the transition from Latin to the individual Romance languages.

I intend to fill this gap to a certain degree by addressing the peculiar evolution of the two patterns in Romanian. Therefore, in the present paper, I aim to answer the following research questions: (i) how did the two constructions evolve during the period starting with the first Romanian written text until the present day?; could this evolution be described as an instance of constructionalization, or do we witness a series of constructional changes happening within the same constructional node?; and (ii) how do changes in the inventory of the state nouns occurring in these constructions in Romanian affect the evolution and the productivity of the constructions themselves?

In order to provide an answer to these research questions, I will test by means of empirical studies the following hypotheses. With respect to the first research question, my hypothesis is that the evolution of the two constructions in Romanian represents a case of constructionalization, with the MIHI EST as a new construction conveying a psychological and a physiological meaning and the HABEO construction becoming typical for abstract possession. As for the second research question, I hypothesize that the investigated state nouns first occurred in the HABEO construction with an experiential meaning and have shifted over time to the experiential MIHI EST construction, which has become more appropriate for expressing physiological and psychological states with a temporary nuance. Such an evolution also generates the expectation that these specific state nouns occurred more frequently with the HABEO experiential construction in old Romanian than in the present-day language.

2. Theoretical Approach

In order to answer the research questions stated above, I adopt a constructionist, usage-based approach, in which I combine elements of the Construction Grammar framework (Langacker 1987, 2003; Fillmore 1996; Croft 2000; Croft and Cruse 2004; Goldberg 2006; Hoffmann 2013), with elements pertaining to the Diachronic Construction Grammar framework (Fried 2009; Traugott and Trousdale 2013; Barðdal et al. 2015). As for the competition between the two constructions in Romanian, I will make use of a model of language change based on the concept of functional competition (De Smet et al. 2018, p. 198).

2.1. Construction Grammar

The two constructions under scrutiny are conceived of as constructions in the sense described in Construction Grammar. In this framework (cf. in particular Goldberg 1995, 2006, 2013; Croft 2001), constructions are defined as conventionalized associations of form and function: [[Form] ↔ [Meaning]]. Such associations are unique and conventionalized and show different degrees of internal complexity and lexical specificity. In addition to words (e.g., *book*), they include, among other things, prefixes (e.g., *pre-*) and idioms (e.g., *get one's ducks in a row*).

As opposed to the modular approach of Generative Grammar, in which the arbitrary character of language is entrusted to the lexicon, while syntax consists of the organization of lexical units, Construction Grammar also acknowledges syntactic constructions to be pairings of *form* and *function* (e.g., *John gives/slides the book to Mary*). For example, the ditransitive construction involves a subject, a verb, and two objects on the formal side and evokes a transfer of an entity by an agent to a beneficiary on the functional side (Goldberg 2006, p. 5). Due to the association of this meaning with the form, it is possible for verbs like *slide*, which do not intrinsically evoke a transfer, to express, nonetheless, such a scenario, and thus to behave like *give*, a verb denoting a transfer through its lexical semantics (Goldberg 2006, p. 7). Constructions are seen as the minimal units of language.

Their inventory in a given language is referred to as the “constructicon”, by analogy to the term “lexicon” (Goldberg 2019, pp. 35–37).

2.2. Grammaticalization, Constructionalization, and Constructional Change

Originally, grammaticalization denoted a process through which a lexical item changes over time into a morphological item (Meillet 1912). Although it is a relatively old concept with a long history, especially in the field of morphology (Meillet 1912), the concept of grammaticalization has gained an “expanded” meaning during the last half of the century, being mentioned in nearly all of the studies involving language change and identified with the processes that create such changes, as well as with the theory modeling these changes itself (cf., among others, Kuryłowicz 1965, p. 69; Lehmann [1982] 2015, 2002, p. 7; Burling 1992, p. 300; Traugott and Trousdale 2013; Hilpert 2013, 2018; Barðdal and Gildea 2015; Gildea and Barðdal 2023). This controversial “expanded” concept of grammaticalization has more recently been linked with the framework of Diachronic Construction Grammar, which is seen as its “well-elaborated agreed-upon theory” (Gildea and Barðdal 2023, p. 2; Noël 2007; Gisborne and Patten 2011; Hilpert 2013, 2018; Coussé et al. 2018). Indeed, terms such as *lexicalization*, *constructionalization*, or even *pragmaticalization* are often defined by invoking a specific flavor of the well-known concept of grammaticalization.

Constructionalization is a more recent concept. It was first used by Rostila (2004) and then by Noël (2007). It is used to refer to “the development through which certain structural patterns acquire their own meanings, so that they add meaning to the lexical elements occurring in them” (Noël 2007, p. 192), as is the case with the *way* construction (*The wounded soldiers limped their way across the field*, Noël 2007, p. 187). Traugott and Trousdale (2013, p. 22) are stricter in defining this concept. For them, constructionalization is defined as the creation of a form–meaning pairing representing a new type node with its own syntax or morphology and also its own (new) coded meanings, a pairing that has been replicated across a network of language users. Everything that happens with the construction before and after the creation of this new node in the network (before and after the constructionalization) is described as effects of pre-constructionalization and post-constructionalization, which are basically constructional changes.

Constructional changes, as defined by the same scholars (Traugott and Trousdale 2013, p. 22), represent changes affecting one internal dimension of a construction, and they do not involve the creation of a new node in the network. Importantly, these changes in contextual use can be observed prior to and following the constructionalization of a pattern (cf. Traugott and Trousdale 2013, pp. 22–26; Hilpert 2015, pp. 135–36).

2.3. Functional Competition

To account for the competition between the HABEO and the MIHI EST constructions, especially after they have been passed on from Latin to Romanian, I will make use of the concept of *functional competition*. In linguistics, the concept of functional competition is a commonly used metaphor to refer to the “selection struggle” among alternative forms encountered by the speaker during language production (Fonteyn 2019, p. 53; Berg 2014, p. 344). The strength of this struggle depends on the degree of formal or functional similarity between the alternatives, as follows: the higher their similarity, the stronger their competition (Berg 2014, p. 344). Competition is the strongest when the alternatives combine formal and functional similarities (Berg 2014, p. 344).

Once two forms enter into competition with one another over some functional domain, two main scenarios can be distinguished (cf. De Smet et al. 2018, p. 198). Either one of the alternatives is favored, causing its competitor to decline and even disappear (Leech 2009), or each of them subsists but specializes in distinct functional domains (Cacoullos and Walker 2009). The outcomes of these scenarios, called respectively *substitution* and *differentiation* by De Smet et al. (2018, p. 198), have a similar effect on the language system: they increase its degree of isomorphism (De Smet et al. 2018, pp. 198–99). Isomorphism refers to the situation in which there is a functionally motivated division of labor between the forms of

a given language, in line with the more general idea that “a difference in syntactic form always spells out a difference in meaning” (Bolinger 1968, p. 127). Languages are claimed to naturally develop an increasing degree of isomorphism, ideally providing their users with one form for one function (al domain).

In the present paper, I will look into the possibility that (i) the two experiencer constructions under scrutiny may have evolved in Romanian into a new node in the constructional network, hence they may have constructionalized; or (ii) that they may instantiate cases of constructional change, limited to transformations affecting one internal dimension of the construction. The empirical data analyzed in the present study also aim to provide additional evidence in favor of the claim put forward by Vangaever and Illoaia (2021), namely that the path of the functional competition between the two constructions settles in Romanian in terms of differentiation and not in terms of substitution.

3. The Evolution of the MIHI EST and HABEO Patterns

3.1. The Journey of the MIHI EST and HABEO Patterns from Latin to Romance

In Latin, prototypical possession is realized in several ways. Prototypical possession is defined as a relation of ownership between an animate possessor and a concrete possessee that can, in some way, be used by the possessor (Bolkestein 2001, p. 269). Among the several strategies in Latin, the following two cross-linguistically recurrent patterns can be distinguished: a transitive pattern, in which a verb of possession occurs with a (potentially pro-dropped) nominative NP and with an accusative one (3), namely the HABEO pattern; and an intransitive pattern, in which the verb ESSE ‘to be’ is combined with a nominative NP and a dative one (4), more specifically, the MIHI EST pattern (Bolkestein 2001, p. 269).

- | | | |
|-----|--|----------------------|
| (3) | <i>Librum habeo</i> book.ACC have ‘I have a book.’ | (Possessor HABEO) |
| (4) | <i>Mihi est liber</i> I.DAT is book.NOM ‘I have a book.’ | (Possessor MIHI EST) |

In these structures, the nominative NP in the transitive pattern and the dative NP in the intransitive pattern encode the possessor, while the accusative and the nominative NPs, respectively, encode the possessee. Although the possessee is mostly concrete in Latin, it may sometimes be abstract, as in (5) and (6), respectively.

- | | | |
|-----|---|------------------------|
| (5) | <i>Febrim habeo</i> fever.ACC have ‘I have a fever.’ | (Experiencer HABEO) |
| (6) | <i>Mihi febris est</i> I.DAT fever.NOM is ‘I have a fever.’ | (Experiencer MIHI EST) |

These abstract NPs encode the possessee, and hence the transitive and intransitive patterns in (5) and (6) do not express a relation of prototypical possession, but an experience (Bolkestein 1983, pp. 83–84; 2001, p. 269; Fedriani 2011, p. 310; Pinkster 2015, p. 108; Danesi and Barödal 2018, p. 23). The two patterns may be labeled *possessor* HABEO and *possessor* MIHI EST when they occur with a concrete noun, as in (3) and (4) above, and *experiencer* HABEO and *experiencer* MIHI EST when they occur with an abstract noun, as in (5) and (6).

The examples above illustrate that *possessor* HABEO (3) and *MIHI EST* (4) have distinct formal, but similar, functional properties and that the same holds for *experiencer* HABEO (5) and *MIHI EST* (6). Since constructions with similar functional properties often enter into competition, it comes as no surprise that competition exists between HABEO and MIHI EST in both *possessor* and *experiencer* constructions (Baldi and Nuti 2010, §2; Fedriani 2011, pp. 310–11). This competition exists in Latin from the earliest texts onward and might even be inherited from an earlier stage of the Indo-European language (Fedriani 2011, p. 311).

However, as shown by Vangaever and Ilioia (2021), the competition between HABEO and MIHI EST is not absolute (cf. also Fedriani 2011, pp. 310–11). More precisely, in Latin, a HABEO construction can always be replaced by a MIHI EST construction, while the opposite substitution does not hold. This imbalance between HABEO and MIHI EST patterns is argued to be due to a difference regarding the nouns that occur in possessor and experiencer constructions, as follows: the variety of the concrete nouns coding the possessee in possessor constructions is open with both HABEO and MIHI EST, while the inventory of the abstract nouns filling the stimulus slot in experiencer constructions is restricted with both verbs, though more significantly with HABEO (Baldi and Nuti 2010, pp. 260–61; Fedriani 2011, p. 311). This shows that MIHI EST is more specialized in experiencer contexts than HABEO. In the oldest texts, e.g., those of Plautus, experiencer MIHI EST is, moreover, preferred over HABEO, which is reflected in the higher frequency estimated by Baldi and Nuti (2010, pp. 260–61) at 37 experiencer MIHI EST vs. 7 experiencer HABEO examples. However, instances of experiencer HABEO have already been attested in the earliest period (Fedriani 2011, pp. 311–12), contrary to what is claimed by Löfstedt (1963, pp. 76–78). In spite of the lack of quantitatively oriented corpus studies for Classical Latin, it has been argued that, from the 1st century BC onward, experiencer uses of HABEO, as shown in (7), gain in frequency, gradually putting an end to the absolute dominance of MIHI EST (Fedriani 2011, pp. 311–12; Löfstedt 1963).

| | | | | |
|---------------|--------------|---------------|------------|------------------|
| (7) <i>Si</i> | <i>cui</i> | <i>venae</i> | <i>sic</i> | <i>moventur,</i> |
| if | REL.DAT | veins.NOM | so | tremble |
| <i>is</i> | <i>habet</i> | <i>febrim</i> | | |
| he | has | fever.ACC | | |

‘If his veins tremble in this way, he has a fever.’ (Cic. De fato 15)

Löfstedt (1963; see also Fedriani 2011, pp. 310–11) argues that the increasing use of HABEO instead of MIHI EST in experiencer constructions is due to a gradual loosening of the original constraints on the object of HABEO, i.e., that it should be a concrete noun denoting an object that can somehow be used by the possessor. Hence, in Classical Latin, speakers came to gradually exploit the use of HABEO in experiencer contexts, which led to a more intense selection struggle between this construction and experiencer MIHI EST.

According to Fedriani (2011, p. 310), the transition from concrete to abstract possession (or experience) can be analyzed within the context of the “ideas are objects metaphor.” This metaphor conceives of the mind as a container in which abstract entities, such as feelings and emotions, are stored as objects (Kövecses 2003, p. 89). This is in line with the more general idea that the use of abstract nouns consists of “a linguistic technique that allows actions and processes to be treated as if they were things” (Seiler 1983, p. 52).

In Late Latin, HABEO and MIHI EST still occur in both possessor and experiencer constructions. At this stage of the language, however, the competition between HABEO and MIHI EST increases in intensity, and, more significantly, starts settling in favor of HABEO, both in possessor and in experiencer constructions, moving towards total replacement of MIHI EST by HABEO in the Romance languages (Fedriani 2011, p. 311). Knowing that the two main outcomes of functional competition are substitution and differentiation, the generalization of HABEO at the cost of MIHI EST should be analyzed, at this point, in terms of “constructional substitution” (Fedriani 2011, p. 311). The empirical evidence supporting this evolution comes from the Romance languages themselves. While MIHI EST is entirely lost in most of these languages, HABEO can still take as its object both concrete (8a–c) and abstract (9a–c) nouns, and thus occurs in possessor, as well as experiencer, constructions (Stolz et al. 2008; Van Peteghem 2017).

| | | | | | |
|--------|---------------------|--------------|---------------|--------------|-----------|
| (8) a. | <i>Pierre</i> | <i>a</i> | <i>un</i> | <i>livre</i> | (French) |
| b. | <i>Pedro</i> | <i>tiene</i> | <i>un</i> | <i>libro</i> | (Spanish) |
| c. | <i>Pietro</i> | <i>ha</i> | <i>un</i> | <i>libro</i> | (Italian) |
| | Peter | has | a | book | |
| | 'Peter has a book.' | | | | |
| (9) a. | <i>Pierre</i> | <i>a</i> | <i>faim</i> | | (French) |
| b. | <i>Pedro</i> | <i>tiene</i> | <i>hambre</i> | | (Spanish) |
| c. | <i>Pietro</i> | <i>ha</i> | <i>fame</i> | | (Italian) |
| | Peter | has | hunger | | |
| | 'Peter is hungry.' | | | | |

Vangaever and Ilioiaia (2021, p. 254) raise the question of why the competition between HABEO and MIHI EST settled in the Romance languages in terms of substitution and not in terms of differentiation. One possible reason could be the perception, already existing in Latin, of HABEO being a more expressive construction (Fedriani 2011, pp. 312–13). For instance, grammarian Donatus explicitly testifies to the expressivity of HABEO in his comment on verse 40 of Terentius' *Andria*, the 4th century AD, as follows: *Plus dixit "in memoria habeo" quam si dixisset "scio"* ("I have in memory" means more than "I know"). As a second reason, the scholars invoke a more general tendency observed in the evolution of the Indo-European languages toward a more transitive syntax (cf. Bauer 1993, p. 65). The two scholars tend to consider the syntactic pressure exerted by the spread of transitivity as having a more significant role in this substitution. However, since the replacement of the intransitive MIHI EST pattern by the transitive HABEO pattern may be considered as part of a large-scale Indo-European syntactic drift, one might raise the question of why it did not take place in the transition from Latin to Romanian.

3.2. Exploring the Evolution of the Competing Constructions MIHI EST and HABEO in Romanian

Indeed, contrary to the other Romance languages, Romanian has preserved both HABEO and MIHI EST, in possessor, as well as experiencer, contexts (10) and (11) (Vangaever and Ilioiaia 2021; Niculescu 2013, pp. 185–86). It has to be noted that the use of the MIHI EST pattern in possessor contexts is restricted to identificational clauses (11a). The nominative NP is the predicate of another, potentially pro-dropped, nominative NP (e.g., *Ioana*, in this example) acting as the subject of *fi* 'to be' and being referentially identified by the "property-denoting nature of the possessee NP" (Niculescu 2013, p. 186).

| | | | | |
|---------|------------------------------|--------------|--------------|-------------------|
| (10) a. | <i>Am</i> | <i>un</i> | <i>fiu</i> | |
| | have | a | son.ACC | |
| | 'I have a son.' | | | |
| b. | <i>Am</i> | <i>frică</i> | | |
| | have | fear.ACC | | |
| | 'I am afraid.' | | | |
| (11) a. | <i>(Ioana)</i> | <i>Îmi</i> | <i>este</i> | <i>cumnată</i> |
| | Ioana.NOM | me.DAT | is | sister-in-law.NOM |
| | 'Ioana is my sister-in-law.' | | | |
| b. | <i>Mi-</i> | <i>e</i> | <i>frică</i> | |
| | me.DAT- | is | fear.NOM | |
| | 'I am afraid.' | | | |

The survival of HABEO and MIHI EST in the two constructions excludes the possibility that their competition in the transition from Latin to Romanian settled in terms of substitution. Given that the other main outcome of functional competition is differentiation, it is thus expected that HABEO and MIHI EST specialized in different possessor and experiencer contexts.

In their paper, Vangaever and Ilioiaia (2021) investigate how this competition has evolved in Romanian from the beginning of the 16th century (with the first Romanian text dating from 1521) to the present-day language via a corpus study. Through this study, the two scholars verified the following two main hypotheses: (i) the Possessor Experiencer Differentiation Hypothesis and (ii) the Experiencer Differentiation Hypothesis. By invoking

the Possessor Experiencer Differentiation Hypothesis, they assumed that the competition between the two constructions evolves toward a neat division of labor, in which HABEO becomes restricted to possessor constructions and MIHI EST to experiencer constructions. By invoking the Experiencer Differentiation Hypothesis, the two scholars expected that HABEO is preserved in experiencer constructions and that it specializes in contexts from which experiencer MIHI EST is excluded.

The results of the corpus study carried out by Vangaever and Ilioia (2021) contradict the first hypothesis, while the second hypothesis is confirmed. More specifically, the examined data show that, in Romanian, HABEO does not become restricted to possessor constructions, but occurs also in experiencer constructions, where it shares an increasing inventory of abstract nouns with experiencer MIHI EST. As for the second hypothesis, the data show that the competition between the two constructions, which takes place this time within the experiencer domain, settles in terms of a specialization in distinct sets of abstract nouns. Indeed, experiencer MIHI EST occurs in present-day Romanian increasingly with abstract nouns expressing psychological and physiological states. To the contrary, experiencer HABEO, which, in Old Romanian, occurred with a larger inventory of nouns denoting psychological and physiological states, occurs in the present-day language with a diversity of abstract nouns denoting various states, but to a much lesser extent with nouns expressing psychological and physiological states. Nevertheless, as pointed out by a reviewer, HABEO is still well-formed with psychological or physiological states in generic, presentational contexts, especially when there is further modification of the noun denoting a state, as, for example, in *Ion are o foame de lup* 'Ion has a hunger of wolf' (meaning that he tends to eat a lot), whereas the experiencer MIHI EST seems to rather convey a temporary nuance to the state expressed.

As is well known, the lexical variation in a specific slot of a syntactic construction is an important parameter to measure the productivity of a construction, i.e., its degree of schematicity and, from the point of view of the language user, its applicability. The more lexical variation there is, the higher the degree of schematicity it has, and, hence, the more productive the construction is (Barðdal 2008, p. 22). The lexical variation that characterizes the set of abstract nouns occurring in the experiencer HABEO construction may thus be linked with a higher degree of productivity, as compared to the experiencer MIHI EST construction, which is limited to nouns expressing psychological and physiological states. In contrast, the low lexical variation (fewer types) of the set of nouns occurring with the experiencer MIHI EST, in combination with a high semantic coherence, indicates a lower degree of schematicity, and, consequently, a lower position on the productivity cline (Barðdal 2008, p. 38). The schematicity of a pattern is the result of the process of entrenchment, which is defined as the continuous reorganization of linguistic knowledge caused by the repeated usage activities in usage events (cf. Barðdal 2008). The more entrenched and schematic a pattern is, the higher the chance of it evolving into a new node in the network, i.e., as a new construction with its own meaning.

Indeed, Ilioia (2020, 2021), Ilioia and Van Peteghem (2021), and Niculescu (2013, p. 186) observed that the MIHI EST pattern is mostly associated with nouns from the field of psychological and physiological states in Romanian, while, for HABEO, no semantic categories have been proposed. In line with the Experiencer Differentiation Hypothesis, Vangaever and Ilioia (2021) show that, over time, the nouns used with HABEO tend to be different from those occurring with MIHI EST. The empirical data on Romanian analyzed by the two scholars indeed show that the use of HABEO with psychological and physiological stimuli gradually decreases from the 17th century until today, with such configurations being perceived as ill-formed in the present-day language (e.g., *? am frică* 'I'm afraid').

Bearing this in mind, the question arises as to how the two constructions interact with the set of nouns with which they occur throughout the centuries. Does MIHI EST occur with nouns other than those selected by the HABEO construction? And, if this is the case, can this evolution be described in terms of constructionalization? Does MIHI EST receive a degree of

schematization in Romanian that allows it to be considered as a new node in the network of the construction, as opposed to the situation in Latin?

These issues will be tackled in the next sections, where I first investigate, from a diachronic perspective, the dynamics in the evolution of a select set of nouns occurring in the two constructions. Then, I will address the arguments in favor of an analysis in terms of constructionalization of the MIHI EST in Romanian.

4. Empirical Case Studies

4.1. Methodology

In order to understand the dynamics of the set of nouns that occur with the MIHI EST and HABEO constructions, I carried out two corpus studies based on texts from pre-21st century Romanian and the present-day language. For pre-21st century Romanian, I worked with a corpus made by myself, which is accessible on demand for research purposes on the Sketch Engine platform. This corpus, labeled *Pre-21st century Romanian*, contains nearly six million words. As for the present-day language, I worked with the *Romanian Web 2016 (roTenTen16)* corpus containing over two billion words, which was compiled and made available on Sketch Engine. It has to be specified that the collection of data was a two-step process, as follows: the detection of the nouns occurring with the MIHI EST construction in Romanian was a necessary step in order to reveal all of the other experiencer structures in which these nouns occur, and to avoid, at the same time, an enormous amount of noise coming from the frequency and flexibility of the verb *fi* 'be' on the one hand, and that of the nouns that combine with it on the other hand.

Using the advanced query option on Sketch Engine, all examples containing the verb *fi* 'be', preceded by a dative clitic and followed by a noun, were gathered from the corpus for present-day Romanian, roTenTen16, provided by Sketch Engine, and the corpus for pre-21st century Romanian, created by myself. By means of an advanced query, I searched for all structures of the type [DAT *fi* N] (cf. *mi-e lene* 'I am lazy'), where N stands for any noun that can combine with an experiencer in the dative and the verb *fi* 'be'.

The query returned 154,492 examples for present-day Romanian, which I restricted to a random sample of 100,000 examples, the maximum amount of data that can be exported from Sketch Engine. The same query in the corpus for pre-21st century Romanian returned 2,278 examples, which were all preserved.

For each of the two samples, I automatically generated a list of different sequences (of the type *mi-e foame* 'I am hungry' or *li-era foame* 'they were hungry') using the Frequency tool on the Sketch Engine platform. The full list was then exported to Excel and manually annotated. After removing the noise, the remaining examples were further annotated, and a final list was made containing the nouns occurring in this construction in all periods of Romanian. Due to imperfections in tagging or inconsistency in the use of diacritics in the older texts, the query returned a larger amount of noise for pre-21st century Romanian than for present-day Romanian. In spite of that, the data for pre-21st century Romanian allowed me to collect 29 different state nouns occurring in the MIHI EST construction, whereas the corpus for the present-day language yielded 95 different nouns in this construction.

The aim of the second phase of the corpus study was to search for all structures that can host the gathered nouns, besides the MIHI EST construction, both in present-day Romanian and in pre-21st century Romanian. Several structures, such as the HABEO construction, and other verb constructions with a nominative, accusative, or dative experiencer, were found. To achieve the maximum number of each of the selected structures for each noun, and to minimize the amount of noise, I created very specific queries for each of the nouns from the list collected during the first phase of the study. An example of a simplified query is [DAT v N] | [NOM v N] | [ACC v N], where 'v' can be any verb, including *fi* 'be' or *avea* 'have', and N is each time replaced with one of the nouns from the list gathered in the first phase, with a distance of zero to three words between them.

Several peculiarities of the present-day language were considered, such as irregularities in the use of diacritics (cf. *frică* vs. *frica*, *scârbă* vs. *scarbă*/*scârba*/*scarba*/*scîrbă*/*scârba*).

scîrba / *scîrbă* / *scirba*, etc.), or spelling modifications for pragmatic reasons (cf. *frig* vs. *friiig*, or *frică* vs. *fricăăăă* / *ffrică*, etc.). As for the particularities of old Romanian, special attention was given to archaic forms (cf. *hi* in place of *fi* ‘be’, *pohtă* instead of *poftă* ‘craving’, *seate* for *sete* ‘thirst’, etc.) and to inconsistencies in the already mentioned use of diacritics (cf. *supra*). It goes without saying that, when working with old texts, other orthographic peculiarities may occur, which cannot always be predicted (cf. for instance, words broken by brackets). These issues can explain the greater amount of noise that was extracted from the corpus for pre-21st century Romanian.

After considering all of the predictable situations, I ran the query for each noun. From the total number of hits per noun, a sample of 200 sentences was selected, with my goal being to collect approximately 100 relevant examples per noun in order to generate the final dataset. It has to be noted that this was not possible for all of the nouns or for all periods, due to size-related limitations.

By means of these queries, I was able to retrieve 16,550 examples for all periods of Romanian, before eliminating the noise (i.e., not relevant contexts, doubles, or typesetting errors). These examples were extracted and saved in an Excel document. After manually filtering the noise, as well as the examples containing constructions other than HABEO and MIHI EST, my sample for the present study contains 7,007 examples to be analyzed (tokens), among which 3,969 examples (57%) are instantiations of the MIHI EST construction, while 3,038 examples (43%) contain the HABEO construction. As expected, the relevant examples from present-day Romanian are more numerous than those from pre-21st century Romanian. More precisely, 73% (5,094) of the examples represent present-day Romanian, and 27% (1,913) pre-21st century Romanian.

With respect to the gathered set of nouns occurring in the two constructions during both present-day and pre-21st century Romanian, the majority of them are old nouns, either inherited from Latin or loaned from Old Slavic. A few of them are eliminated from the MIHI EST construction in the 19th century (*nevoie* ‘need’) or the 21st century (*jind* ‘craving, longing’)—while they are still used in present-day Romanian in other constructions, including the HABEO construction. New nouns are recruited in every century, mostly first by the HABEO construction, and later on by the MIHI EST construction, as follows: *dor* ‘longing’ (the 16th century: HABEO, end of the 17th century: MIHI EST), *teamă* ‘fear’ (end of the 16th century: HABEO, the 19th century: MIHI EST), and *jenă* ‘embarrassment’ (the 19th century: HABEO, the 21st century: MIHI EST), among others. Some of the newly recruited nouns are loans from modern Romance or surrounding Slavic languages, and others are derivations of already existing, inherited nouns.

Remarkably, starting with 20th-century Romanian, the MIHI EST construction freely allows the occurrence of synonymous nouns in it (cf. Ilioiaia 2021). For instance, the Turkish *zor* ‘hastiness/rush’, a synonym of the derived *grabă* ‘hastiness/rush’, occurs in the dative experiencer construction starting with the 20th century, and is still in use today, as opposed to *grabă*, which was recruited much earlier, and disappeared from it in the 21st century.

Furthermore, more recently, the MIHI EST construction, as opposed to the HABEO construction, has shown a tendency to admit nouns from different semantic fields, such as event (*plecare* ‘departure’), time (*atât de primăvară* ‘so springtime’), and communication (*cuoânt nerostit* ‘untold word’), among others, as pointed out by Ilioiaia (2020). These new nouns are used metaphorically in the MIHI EST construction and are being ‘forced’ into the construction’s meaning, that of a physiological or a psychological state, as shown in (12) and (13). This process is called *coercion* and is defined in the literature as the reinterpretation of a lexical item triggered by the conflict between its meaning and that of the construction in which it is used.

- (12) *Mi- e atât de primăvară încât*
 me.DAT= is so of springtime that
șoptesc întunericului să [...]
 whisper.1SG dark.the.DAT SUBJ [...]
 ‘I feel so springtime that I whisper the darkness to (...)’
 (facebook.com/permalink.php, accessed on 5 May 2018)
- (13) *mi- e cuvânt nerostit...*
 me.DAT= is word untold
 ‘I feel in a way that I cannot express (...)’
 (alexsmallthings-desprenimic.blogspot.com, accessed on 5 May 2018)

Recall that, as Lauwers and Willems (2011, p. 1220) highlighted, in Construction Grammar, *coercion* constitutes a major argument in favor of the existence of constructions as independent form/meaning pairings. The capability of a construction to change the meaning of a lexical item that occurs in it has been used as an empirical test to argue in favor of the construction status of structures carrying a particular meaning on their own, irrespective of the lexical items that instantiate them.

4.2. Results

4.2.1. Distribution of the Inventory of the State Nouns under Investigation

Romanian, like most other languages, has several possibilities to express psychological or physiological states by means of a structure of the type [EXPERIENCER V NOUN_{STATE}]. As mentioned in the previous section, besides the HABEO and the MIHI EST constructions, there are other experiencer constructions that occur with the state nouns under scrutiny (14)–(16). Among all of the possible competing constructions, there are some that select for a dative (14), accusative (15), or nominative (16) experiencer.

- (14) *Mi s- a făcut dor*
 me.DAT SE= has made longing
de părinți
 of parents
 ‘I began to miss my parents.’
 (1882, Petre Ispirescu, *Basme*)
- (15) *M- a cuprins frica*
 me.ACC= has sized fear.NOM=ACC
 ‘I became afraid.’
- (16) *Băiatul prinse frică de cățel*
 boy.NOM caught fear of puppy
 ‘The boy became afraid of the puppy.’
- (17) *Mariei i- e frică dar*
 Maria.DAT her.DAT= is fear but
nu vrea să zică
 not wants SUBJ say.SUBJ.3SG
 ‘Maria is afraid but she doesn’t want to say it.’
- (18) *După acéia, Gligorie- vodă, având*
 after that Gligorie- voivode having
grijă și mare frică, [...]
 worry and big fear [...]
 ‘Then, since Grigorie-voivode was worried and very afraid, [...].’
 (1659, Constantin Cantacuzino, Letopisețul)

However, based on my data, the frequency of these other syntactic patterns is not as high as the frequency of the HABEO construction (3,038 occurrences) or the MIHI EST construction (3,969 occurrences), as shown in (17) and (18). This can be seen in Figure 1 below, where *Other* refers to the other types of constructions and corresponds to the remaining 1,451 occurrences from the total of 8,458 relevant examples in my dataset, after excluding those containing the HABEO or the MIHI EST construction.

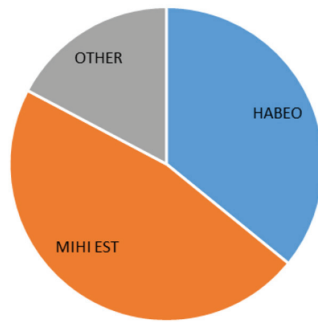


Figure 1. Distribution of experiencer constructions in my dataset (cf. Figure 5.3 by Ilioiaia 2021).

Therefore, in what follows, I focus only on the two most frequent constructions, the HABEO and the MIHI EST experiencer constructions. I examine how the set of state nouns that occur with them evolves throughout the centuries, aiming to make some connections between the changes in the inventory of nouns and the evolution and productivity of the constructions themselves.

4.2.2. Dynamics of the Set of Nouns Occurring in the HABEO and MIHI EST Constructions

Remarkably, a first glance at the data reveals that, more recently, not all of the nouns that occur with the MIHI EST experiencer construction also occur with the HABEO experiencer construction, a fact that seems to argue against the hypothesis that the state nouns occurring in the MIHI EST construction first occurred in the HABEO construction in Romanian. In order to understand this behavior, it is important to highlight that, starting with the 19th century, the MIHI EST construction expands its semantic domain. Basically, as pointed out by Ilioiaia (2020, 2021), the MIHI EST construction becomes open to synonymous nouns and to other nouns not only from the semantic field of psychological and physiological states, but also from different semantic fields, such as time (*noapte* ‘night’, *toamnă* ‘autumn’), nature (*foc* ‘fire’), weather (*ger* ‘freezing’, *furtună* ‘storm’), matter (*vomă* ‘vomit’), events (*vis* ‘dream’), and act/object (*joc de copii* ‘game of children’), among others. When part of the MIHI EST construction, these new nouns, most of which are used metaphorically, express a physiological or a psychological state, as shown in the examples in (19) and (20) below and in (22) further below.

(19) *Îmi* *este* *durere* *și* *-mi*
 me.DAT= is pain and =me.DAT
este *furtună*
 is storm
 ‘I feel pain and I feel tormented (in my soul)’

(reteualiterara.ning.com/m/blogpost?id=1971741, accessed on 5 May 2018)

(20) *Mi-* *e* *joc* *de* *copii*
 me.DAT= is game of children
 ‘I feel playful’

(alexsmallthings-desprenimic.blogspot.com, accessed on 5 May 2018)

Not surprisingly, the nouns that recently became compatible with the MIHI EST construction, as well as the newly entered state nouns, are all nouns with a very low frequency, mostly up to three or four occurrences in the construction, i.e., *hapax legomena* (nouns that occur only once in my dataset in the construction under scrutiny), *dis legomena* (nouns occurring twice), or *tris legomena* (nouns occurring three times). Importantly, when combined with the HABEO construction, these new nouns do not convey a psychological or a physiological meaning. Nevertheless, due to their common character, these nouns inevitably occur with HABEO in other types of constructions, mostly conveying an abstract possession meaning or a totally different meaning not relevant here. Take, for instance, the

noun *toamnă* ‘autumn’, a very common noun that does not belong to the semantic field of psychological or physiological states. This noun conveys an abstract possession meaning when combined with a HABEO construction (21) but, when used metaphorically in the MIHI EST construction, as shown in (22), it expresses a psychological state.

- (21) *Anul trecut am avut o*
 year.the past have.1SG had an
toamnă ploioasă
 autumn rainy
 ‘Last year I had/ we had a rainy autumn.’
- (22) *Mi-e toamnă, iubito!*
 me.DAT= is autumn.NOM beloved.VOC
 ‘I feel melancholic, my love!’

(intelepciune.ro, accessed on 5 May 2018)

In light of the ability of the MIHI EST construction to coerce nouns coming from other semantic fields into its own meaning, i.e., that of a psychological or a physiological state, the data presented in this section provide valuable evidence in favor of the construction_{new} status of the MIHI EST structure, hence, in favor of an evolution in terms of constructionalization.

4.2.3. Experiencer HABEO vs. Experiencer MIHI EST

In what follows, I first give an overview of the competition between the two experiencer constructions, based on the entire dataset. Subsequently, I describe a few diachronic case studies in order to investigate per noun if a shift takes place from the HABEO to the MIHI EST construction.

Figure 2 shows the competition between the two constructions when used with the state nouns in my dataset, for all periods combined. Note that, for readability purposes, the graph plots only the nouns with a relative frequency of above 0.2%. The frequencies shown in this graph have been obtained by dividing the absolute frequency of each noun by the total number of occurrences in the dataset, i.e., 7,007 occurrences. From this graph, it can be observed that the nouns occurring less frequently in the MIHI EST construction still have a higher frequency in the HABEO construction, whereas the more frequently a noun is used in the MIHI EST construction, the less frequently it occurs in the HABEO construction (cf. also Ilioiaia 2021, p. 145).

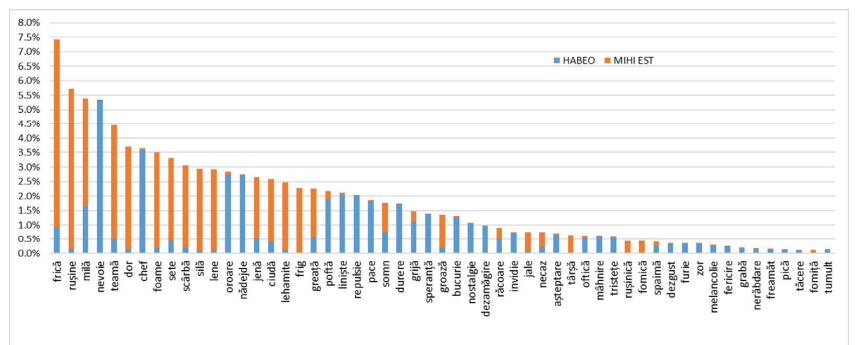


Figure 2. Competition of HABEO vs. MIHI EST—frequent nouns (16th–21st) (cf. Figure 5.4 by Ilioiaia 2021).

As for the evolution of this rivalry, Table 1 offers an overview of the competition between the two constructions throughout the centuries when they occur with nouns denoting a state. Since the sets of examples for each historical period are not of equal size, the absolute figures are followed by relative figures in the table. The relative figures were

calculated by dividing the absolute figures by the total number of examples in the dataset (7,007). These data are visualized in the graph in Figure 3, which illustrates that it is only in the first historical period that the HABEO construction was dominant. Starting with the 19th century, the occurrence of the abovementioned state nouns with the MIHI EST construction prevails, and this supports the claim that these nouns were first employed in the HABEO construction and then passed on to the MIHI EST construction, judged as more appropriate to express psychological and physiological states.

Table 1. Competition of HABEO vs. MIHI EST—diachronic perspective (cf. Table 6.4, Ilioiaia 2021, p. 145).

| | HABEO | % | MIHI EST | % |
|-----------|-------|-------|----------|-------|
| 16th–18th | 232 | 3.3% | 159 | 2.3% |
| 19th | 195 | 2.8% | 335 | 4.8% |
| 20th | 301 | 4.3% | 691 | 9.9% |
| 21st | 2310 | 33.0% | 2784 | 39.7% |
| Total | 3038 | 43.4% | 3969 | 56.6% |

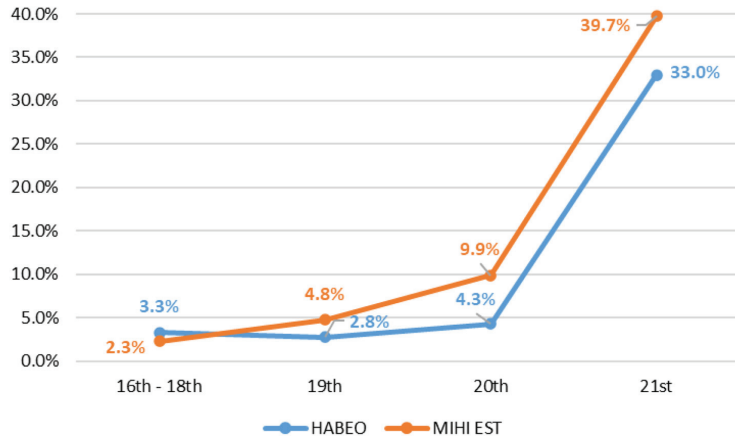


Figure 3. Competition of HABEO vs. MIHI EST—diachronic perspective (cf. Figure 6.5, Ilioiaia 2021, p. 145).

On first sight, the frequency of the nouns occurring in my dataset with the HABEO construction is surprisingly high for the 21st century. However, this may be explained by the high number of new nouns entering the MIHI EST construction, most of which belong to the common vocabulary of the language and, hence, are very frequent. For this period, Ilioiaia (2021, pp. 233–34) shows that 55% of all of the nouns occurring with MIHI EST have a very low frequency in this construction. They are either *dis legomena* (5%) or *hapax legomena* (50%), mostly belonging to other semantic fields, such as events, weather, time, or elements of nature, as illustrated in the graph in Figure 4.

These new nouns already occur with HABEO much more often in my dataset, however, the HABEO configurations into which they enter exclude a psychological or a physiological meaning. This, in fact, supports the hypothesis stated in the introduction of this paper, that the nouns under investigation first occurred in the HABEO construction and shifted over time to the MIHI EST construction, which has become more specialized in expressing psychological and physiological states.

This evolution can also be observed diachronically, as evident from Table 1 above (cf. Ilioiaia 2021), where the nouns denoting a state were more frequently found in the HABEO experiencer construction in the older periods of Romanian than they were in the

present-day language. Indeed, Vangaever and Illoaia (2021) show that, in the older periods of Romanian, HABEO occurred with all kinds of nouns, including concrete, abstract, and even with nouns denoting psychological or physiological states. This situation changes over time, and HABEO becomes typical for expressing predicative possession¹ in Romanian, while its inventory of nouns expressing a psychological or a physiological state decreases considerably, being taken over by the more specialized MIHI EST construction.

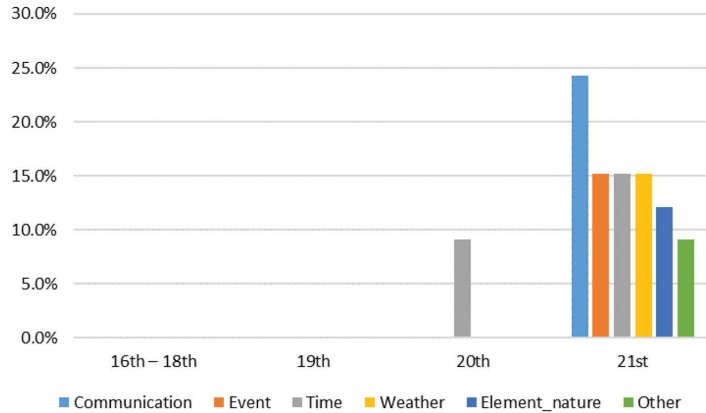


Figure 4. The less frequent semantic classes in the MIHI EST experiencer construction.

The evidence by Vangaever and Illoaia (2021) reinforces the findings of an earlier case study of two of the most frequent nouns entering the MIHI EST construction in Romanian, *frică* and *teamă*, both meaning ‘fear’ (Van Peteghem and Illoaia 2017). This study shows that *frică* was present in both constructions in the 16th century, but was more frequent in the HABEO construction, whereas *teamă* occurred in the HABEO construction at the end of the 16th century, when it was attested in the language for the first time and shifted to the MIHI EST construction later on, in the 19th century.

4.2.4. Case Studies

In what follows, I focus on a limited number of nouns and investigate whether these nouns occurred first in the HABEO construction and then in the MIHI EST construction, and which of the two rival constructions is preferred throughout the centuries. The list of nouns under investigation constitutes a subset of a broader set studied by Illoaia (2021) and contains the following nouns: *frică* ‘fear’, *teamă* ‘fear’, *dor* ‘longing’, *scârbă* ‘disgust’, *poftă* ‘craving’, *milă* ‘pity’, *grijă* ‘worry, care’, *spaimă* ‘fear’, *grabă* ‘rush, hurry’, and *nevoie* ‘need’.

Among these nouns, certain items seem to change completely in terms of their dominant construction throughout the centuries, while others continue to occur in both constructions. When the dominant construction changes, it is mostly from the HABEO to the MIHI EST construction, and only very isolated cases show the opposite direction of change.

A closer look at the first set of nouns, namely those that show an increasing propensity toward the MIHI EST construction after already being attested in the HABEO construction, reveals fascinating details.

For instance, two of the most frequent nouns in the MIHI EST construction, *frică* and *teamă*, both meaning ‘fear’, start by occurring in the HABEO construction, before shifting to the MIHI EST construction (cf. also Van Peteghem and Illoaia 2017). Importantly, these nouns entered into the language at different moments and from different sources, as follows: *frică* (<Gr. *φρίκη*) was first attested in the 15th century, but might be older, whereas *teamă*, derived from *teme* ‘to fear’ (<Lat. *timere*), is first attested in the late 16th century.

As illustrated in Figure 5, *frică*, just like *scârbă* ‘disgust’ (<Sl. *skrŭbŭ*), occurred from the earliest texts in both constructions, but preferred the HABEO construction during the

first historical period (16th–18th), whereas *teamă* and, similarly, *dor* ‘longing’ (<Lat. *dolus*), appear exclusively in the HABEO construction during this period. The situation changes in the 19th century for all of these nouns, when the MIHI EST construction becomes by far the more dominant construction with *frică* and *scărbă*, while *teamă* and *dor* start being increasingly selected by the MIHI EST construction. This first set of nouns may be considered the most advanced in their evolution in the MIHI EST construction, as their use represents the most natural way of expressing these states in Romanian.



Figure 5. Nouns changing their dominant construction—HABEO vs. MIHI EST.

Note that, for the visual representation of the competition between the two constructions in the graphs given in Figures 5–9, I opted for a bar chart, where each bar represents the total occurrences of that specific noun in the two constructions, for the investigated historical period. For instance, in the first graph given in Figure 5, above the first bar shows that, in the period between the 16th and 18th centuries, the noun *frică* ‘fear’ was more frequent in the HABEO construction (blue area) than in the MIHI EST construction (orange area). The subsequent bars in the same graph show that the situation changes in the following centuries and that the noun becomes more frequent in the MIHI EST construction (orange area) than in the HABEO construction (blue area).

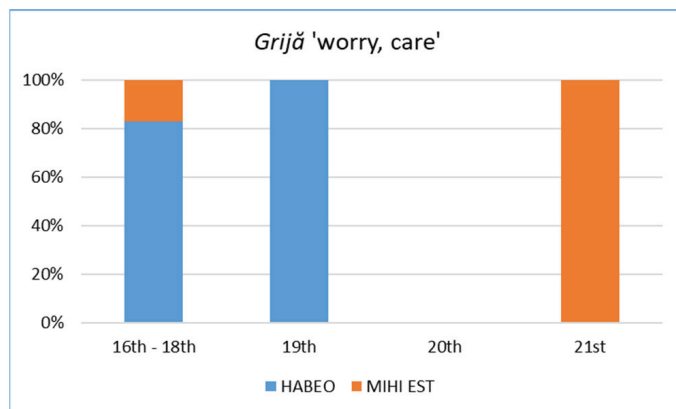


Figure 6. *Grijă* ‘worry, care’—HABEO vs. MIHI EST.

Concerning the last noun in this group, *grijă* ‘worry, care’ (<Bg. *grīža*), it has a particular evolution. After having the HABEO experiencer construction as the dominant construction from the 16th to 18th centuries and as the only selecting construction in the 19th century, it totally disappears from both experiencer constructions in the 20th century, and is revived in the 21st century, where it occurs only in the MIHI EST experiencer construction, as illustrated in (23). My dataset for the 21st century contains 15 such examples. Although some native speakers may find these occurrences rather ill-formed, others do not think so, since the noun is employed in the MIHI EST construction even on official (local) news platforms. It has to be noted that, just as it is the case with other nouns in my dataset, *grijă* ‘worry, care’ is a very common noun in Romanian and does not completely disappear from the language during the 20th century, but only from the experiencer construction under scrutiny. Providing an explanation for this peculiar evolution is not an easy task, since its revival in the MIHI EST construction in the 21st century could be a recent borrowing or a modern calque of the construction as a whole. However, one should not exclude the possibility that its absence during the 20th century is a result of the scarcity of representative data for this specific historical period. In Figure 6, this evolution is visually represented.

- (23) *Dar el a zis aceasta nu pentru că*
 but he have.3SG said this not because
îi era grijă de săraci, ci
 him.DAT was worry of poor.PL but
 ‘But he said this not because he was worried about the poor ones, but...’
 (gorjeanul.ro, accessed on 12 December 2023)

In what follows, a second set of nouns is investigated, containing, among others, *milă* ‘pity’, *poftă* ‘craving’, and *spaimă* ‘fear’. What they have in common is that their frequency in the HABEO construction slightly increases or remains stable throughout the centuries, contrary to what one may expect, given the observed general tendency.

The noun *milă* ‘pity’ (Sl. *milŭ*) was attested very early in the MIHI EST construction (end of the 15th century). Occurrences of *milă* in the HABEO construction are not found in my dataset before the beginning of the 17th century (cf. Figure 7). The path followed by this noun contradicts my hypothesis that the state nouns occur first in the HABEO construction before being recruited by the MIHI EST construction. A closer look at the distribution of its occurrences reveals that *milă* is mostly found in translated texts, as compared to the original texts between the 15th and 16th centuries (11 occurrences in translated texts vs. 6 occurrences in original texts). This could suggest either that this noun was used in a cognate MIHI EST construction in the source language and has been introduced in Romanian through a loan translation of the structure, or that its absence in the HABEO construction represents a consequence of the scarce data for this period of Romanian. However, this does not change the fact that this noun is first attested in the MIHI EST construction, and only a century later in the HABEO construction.

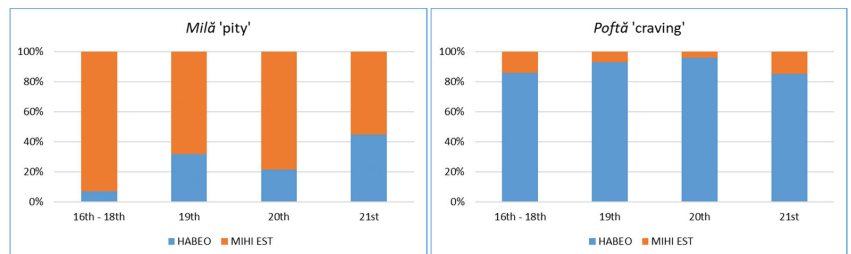


Figure 7. *Milă* ‘pity’ and *poftă* ‘craving’—HABEO vs. MIHI EST.

As for *poftă* ‘craving’, this noun, derived from the verb *pofti* (<Sl. *pohotěti*), shows a particular evolution, since it has a very slow increase in frequency in the MIHI EST

construction, while HABEO continues to be its dominant construction until the present-day language (cf. Figure 7).

The question arises as to whether other nouns may also show this tendency, having a low or decreasing frequency in the MIHI EST construction, while HABEO becomes dominant. A case in point is *spaimă* ‘fear’, which can still be found in the MIHI EST construction in the 21st century, but has a considerably lower frequency, which decreases to half (see Figure 8).

The remaining two case studies refer to *grabă* ‘hurry’ and *nevoie* ‘need’, nouns that disappear completely from the MIHI EST construction in the 21st century and are today only found with the HABEO construction. Their evolution looks similar to the evolution of *spaimă* ‘fear’, with the only difference being that *spaimă* continues to have a reasonable frequency in the present-day language with the MIHI EST construction. It has to be noted though, that these three nouns have never really been deep-rooted in the MIHI EST construction, given that their highest frequency in this construction is six occurrences for *spaimă* in the 19th century and only one occurrence for each of the other two nouns for the same historical period (cf. Figure 9 for *grabă* and *nevoie*).

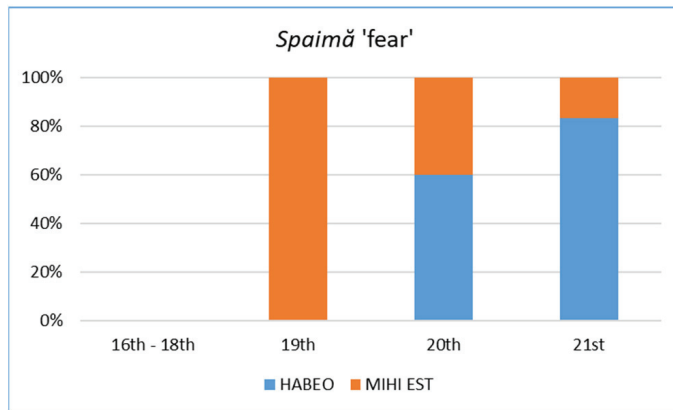


Figure 8. *Spaimă* ‘fear’—HABEO vs. MIHI EST.

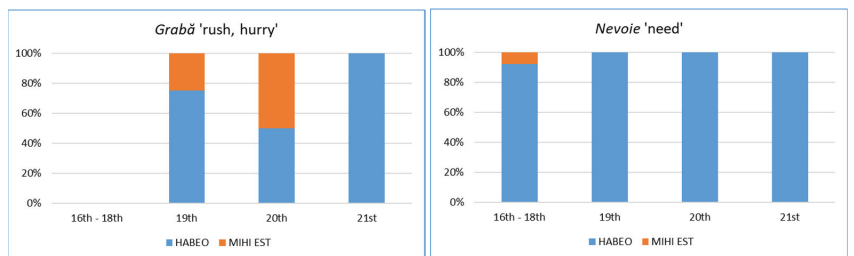


Figure 9. *Grabă* ‘rush, hurry’ and *nevoie* ‘need’—HABEO vs. MIHI EST.

What might the underlying factors be behind the fact that nouns such as *grabă*, *nevoie*, or even *spaimă*, discussed above, take such a turn and do not follow the general trend, that of changing from the HABEO to the MIHI EST construction? A possible explanation is the competition with other synonymous well-established constructions (cf. *spaimă* vs. *frică* ‘fear’) or with newly borrowed elements (cf. *grabă* vs. *zor* ‘rush, hurry’). Yet, in the case of *nevoie*, which occurs in my corpus only once in the MIHI EST experienter construction, in a translation from the 16th century, the explanation could be either that it was an error made by the translator or a literal translation of a structure from the source text, which did not reoccur in the construction in the following centuries. However, one cannot ignore that, while most of the state nouns under scrutiny predominantly occur as bare nouns and

without complementation (cf. *?am/mi-e frică (de tine)* ‘I am afraid (of you)’), *nevoie* seems to require the presence of a complement in order to be well-formed, even in the HABEO construction (cf. *am/?mi-e nevoie *(de tine)* ‘I need (you)’). This last point may uncover another reason why *nevoie* was not perceived by the speakers as being well-formed in the MIHI EST construction but was considered as more compatible with the HABEO construction.

Given the tendencies observed, the question arises as to how many of the most frequent nouns in the 21st century have the HABEO construction as the dominant construction. To obtain such a ranking, I calculate the relative frequencies in present-day Romanian of all of the nouns occurring in the two constructions taken together and sort them by their total (cf. Figure 10). The relative figures were calculated by dividing the absolute frequencies by the total number of examples in the dataset (7007 occurrences).

Seven out of the ten nouns visualized in the graph in Figure 10 have MIHI EST as the dominant construction to express states, whereas three of them prefer the HABEO construction. One of these three nouns is *nevoie* ‘need’, which, as already mentioned, has been found in my dataset only once in the MIHI EST construction, in a translated text in old Romanian, and, hence, may be considered as an accidental occurrence, since it did not develop further in this construction (cf. also Ilioiaia 2021). The other two nouns, *chef* ‘mood’ and *oroare* ‘horror’, are both newly recruited by the MIHI EST construction in the 21st century. I consider that it is too early to predict the evolution of these two nouns, since they only started occurring in the MIHI EST construction in the 21st century, more precisely, *chef* with three occurrences and *oroare* with eight occurrences. Nevertheless, their high frequency in the HABEO construction corroborates my hypothesis that the state nouns under scrutiny first occur in the HABEO construction before shifting to the MIHI EST construction, confirmed by most of the data for the older periods of Romanian.

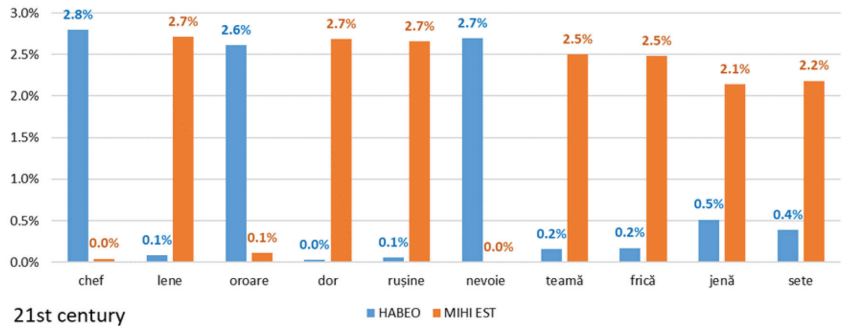


Figure 10. The ten most frequent nouns in the 21st century—HABEO vs. MIHI EST (cf. Figure 5.17 by Ilioiaia 2021).

5. Discussion

Based on the evolution of the state nouns with respect to the two competing constructions, experiencer HABEO and experiencer MIHI EST, presented in the previous section, I suggest that, once a noun has been recruited by the MIHI EST construction, one of the following paths will be taken:

- (i) The state noun decreases in frequency in the HABEO construction (immediately or within a century) or disappears completely from it, while it increases in frequency in the MIHI EST construction. This is the path followed by the following nouns: *frică* ‘fear’, *teamă* ‘fear’, *dor* ‘longing’, *scârbă* ‘disgust’, and *grijă* ‘worry, care’.
- (ii) The state noun keeps occurring in both constructions throughout the centuries, with the HABEO construction still being the dominant one. This is the case for *milă*, ‘pity’, *poftă* ‘craving’, and *spaimă* ‘fear’.
- (iii) The state noun occurs for a period in both constructions but continues to increase in frequency in the HABEO construction while decreasing in frequency in the MIHI

EST construction. This evolution has been observed for *grabă* ‘rush, hurry’, and *nevoie* ‘need’.

In an attempt to provide a theoretical explanation for each of these three paths, I propose to look at the evolution of a filler (i.e., the noun denoting a state in our case) with respect to a specific slot (i.e., the opening or the position of that noun in the argument structure) as corresponding to a certain degree of entrenchment of that specific pattern in the network. The entrenchment is caused by repeated usage and often, as a result, has a high degree of schematization.

This being said, the three paths identified in the evolution of the inventory of nouns in the HABEO and the MIHI EST constructions may be seen as different phases in the process of schematization and, hence, the constructionalization of the constructions. The nouns following the first path are among the most frequent ones and show a high degree of entrenchment in their use with MIHI EST, given their ‘fixation’ in the construction. The second path corresponds to a transition phase, which may result in the entrenchment of the nouns in the MIHI EST construction or may lead to their elimination from this construction and to their ‘fixation’ in the HABEO construction. The third path corresponds to a phase in which the state noun becomes more preferred in the HABEO construction than in the MIHI EST construction, with which it eventually becomes semantically incompatible.

On a more general note, from these case studies, I conclude that most of the state nouns under scrutiny tend to first occur in the HABEO construction and then in the MIHI EST construction, with the only exception being the noun *milă* ‘pity’, which is first attested in my dataset in the MIHI EST construction. For some nouns, the change occurs very abruptly, whereas, for other nouns, it takes centuries. The change is, for some nouns, definitive and in favor of one of the two constructions, whereas other nouns continue to occur in both constructions.

This study also highlights that, although most of the nouns occurred first in the HABEO construction, not all of them had deep roots in this construction when they were recruited by MIHI EST. A possible explanation for this is that these nouns had recently entered into the language. In contrast, some of the nouns entering the MIHI EST construction in the 21st century show a considerably high frequency in the HABEO construction, since they were borrowed into the language one century earlier (*oroare*, in the 20th century), or even two or more centuries earlier (*chef*, in the 19th century). Given this, it is difficult to predict whether the new nouns entering the MIHI EST construction in the 21st century will be able to abandon the HABEO construction, in which they seem rather deep-rooted. One thing is sure, though, their access to the experienter MIHI EST construction has been granted. Note also that the nouns entering the experienter MIHI EST construction in the 21st century and belonging to other semantic fields convey a totally different meaning when they occur in the HABEO construction. I refer here to nouns like *noapte* ‘night’, *furtună* ‘storm’, or *toamnă* ‘autumn’, with some of them illustrated in the examples in (19)–(22) in Section 4.2.2, which, when occurring in the HABEO construction, convey a possessive instead of an experiential meaning. This is a sign that the MIHI EST construction is now able to attract its own nouns, independent of their association with the experienter HABEO construction, and can coerce them into its psychological and physiological meaning.

Let us now look at these findings and interpret them from the perspective of the historical changes through which they may be characterized. One of the research questions stated at the beginning of this paper, was whether the special evolution of the two constructions can be described as a case of constructionalization or constructional change. An answer to this issue remains difficult to provide, given the difficulty in teasing apart the two concepts. Following the definition given by Noël (2007), who considers that we can talk about constructionalization when a structural pattern acquires its own meaning in such way that it adds meaning to the lexical elements occurring in it, there is no doubt that the experienter MIHI EST construction represents a case of constructionalization, since it has the ability to coerce nouns from different semantic fields into its own meaning. Traugott and Trousdale (2013, p. 22), however, propose to qualify the change in a specific construction

as a case of constructionalization only if the historical process results in the creation of a form—meaning pairing representing a new type node, which has been replicated across a network of language users. The difficulty one encounters when attempting to make such a decision comes, as I understand it, from the fact that constructionalization always implies constructional changes, which are said to occur before or/and after the creation of the new node in the network, even though the opposite may not always be true. So, how do we recognize when a construction has met such a degree of schematization that it can be considered as a new node in the network, if not when it has become so specialized that it coerces elements from other semantic fields into its own meaning?

With respect to the HABEO construction, such a degree of schematization is very difficult to demonstrate in Romanian, given how spread the use of HABEO is in several types of constructions. On the contrary, such an evolution is easier to show for the MIHI EST pattern. Based on the analysis of my data and reinforced by the recent tendency toward innovation of the MIHI EST construction, I argue that there is substantial empirical evidence in favor of an analysis of the experiencer MIHI EST construction as a case of constructionalization. The following arguments strengthen this statement:

- (i) The data provide evidence that the MIHI EST construction is, over the centuries, increasingly used with state nouns expressing psychological and physiological states with a temporary nuance, mostly recruited from the experiencer HABEO construction, which shows an opposite tendency, with the state nouns decreasing in frequency. Their evolution provides additional evidence in favor of an analysis in terms of constructional differentiation (Vangaever and Illoaia 2021, following De Smet et al. 2018, p. 198), since experiencer HABEO continues to occur with abstract nouns but specializes in the contexts from which experiencer MIHI EST is excluded.
- (ii) The MIHI EST construction recruits nouns from different semantic fields, and, more importantly, coerces them into its own particular meaning, whether that of a psychological or a physiological state.
- (iii) The Latin experiencer MIHI EST pattern evolves in Romanian into a new, specialized construction, which becomes the most natural way of expressing psychological and physiological temporary states, whereas the experiencer HABEO specializes in conveying the meaning of abstract possession.

Indeed, a clear tendency can be observed throughout the centuries, namely the attraction force exercised by the Romanian MIHI EST construction on state nouns, and especially on the state nouns occurring in the HABEO construction, causing the HABEO construction to become less preferred for expressing psychological or physiological states.

6. Conclusions

The aim of this paper was twofold. First, by connecting the development of the MIHI EST and HABEO patterns throughout Latin with their evolution from Latin to present-day Romanian, this paper intended to provide a clearer view of the evolution of the two patterns and of the historical changes that they have passed through after being inherited by Romanian until the present-day language. Second, by having a closer look at the set of nouns that occur in the two constructions in Romanian, I aimed to understand how the changes in the inventory of the state nouns occurring in these constructions affect the evolution and the productivity of the constructions themselves.

With regard to the evolution of the two patterns in Romanian and to the historical changes that they have experienced throughout the centuries, the outcomes of the corpus studies show that both Latin patterns have been inherited by Romanian with their double nature: possessor and experiencer. The way that these constructions managed to find complementary niches of the language is fascinating. Possessor HABEO continues to occur in Romanian with an unlimited inventory of concrete nouns, while possessor MIHI EST specializes in identificational clauses (cf. *(Ea) mi-e cumnată*, lit. she me.DAT is sister-in-law 'She is my sister-in-law'). As for the experiencer patterns, which are of a great importance in the present paper, experiencer MIHI EST specializes in expressing psychological and

physiological states in Romanian, whereas experiencer HABEO, which occurred during the older periods more often with abstract nouns from the semantic field of psychological and physiological states, becomes less preferred for expressing these states in the present-day language. Nevertheless, HABEO continues to occur with abstract nouns pertaining to other semantic fields. Therefore, in terms of functional competition, the evolution experienced by HABEO and MIHI EST can be described as constructional differentiation, as opposed to the constructional substitution specific for the other Romance languages.

The case studies, intended to reveal any particular dynamics in the inventory of the nouns occurring in the two experiencer constructions, show that, in general, the nouns that occur in the MIHI EST construction first occurred with experiencer HABEO and that they all belong to the field of psychological and physiological states. More recently, however, MIHI EST occurs with nouns from other semantic fields, such as events, time, communication, or weather phenomena, and coerces them into its own meaning, i.e., that of a psychological or physiological state.

The specialization of the MIHI EST construction in expressing exclusively psychological and physiological states with a temporary nuance and its ability to coerce nouns from different semantic fields into the meaning of the construction, in corroboration with the fact that the MIHI EST construction becomes the most natural way of expressing this type of state in present-day Romanian, constitutes valuable evidence in favor of an analysis of this construction in terms of constructionalization.

Funding: This research was funded by Research Foundation-Flanders (FWO Fonds Wetenschappelijk Onderzoek) grant number G0D2516N.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this paper will be made available on the Trolling repository. The corpus for Pre-21st century Romanian is accessible for researchers upon request from the corresponding Author.

Acknowledgments: I wish to thank the two reviewers as well as the editors for their suggestions, which certainly improved the quality of the paper. I would like to also thank my colleague, Hannah Booth for making time to proofreading my article and for her helpful suggestions. All errors or misunderstandings are of the author herself.

Conflicts of Interest: The author declares no conflict of interest.

Note

- ¹ Predicativepossession refers here to an asymmetric (alienable) relation between a possessor and a possessee in which the possessor may control the possessed object (Stassen 2009, p. 11). Note that, in the present paper, we exclusively refer to the *have*-predicative possession type, as opposed to other types of predicative possession, such as the ones in which an oblique, genitive, or topic may be involved as the possessor.

References

- Baldi, Philip, and Andrea Nuti. 2010. Possession in Latin. In *New Perspectives on Historical Latin Syntax*. Edited by Philip Baldi and Pierluigi Cuzzolin. Berlin: Mouton de Gruyter, vol. 3, pp. 239–387.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam and Philadelphia: John Benjamins Publishing, vol. 8.
- Barðdal, Jóhanna, and Spike Gildea. 2015. Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. In *Diachronic Construction Grammar*. Edited by Jóhanna Barðdal, Spike Gildea, Lotte Sommerer and Elena Smirnova. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 1–49.
- Barðdal, Jóhanna, Spike Gildea, Lotte Sommerer, and Elena Smirnova. 2015. *Diachronic construction grammar*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Bauer, Brigitte L. M. 1993. The coalescence of the participle and the gerund/gerundive: An integrated change. In *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*. Amsterdam and Philadelphia: John Benjamins BV, pp. 59–71.
- Bauer, Brigitte L. M. 1996. Residues of non-nominative Syntax in Latin: The MIHI EST construction. *Historische Sprachforschung/Historical Linguistics* 109: 241–56.

- Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.
- Berg, Thomas. 2014. Competition as a unifying concept for the study of language. *The Mental Lexicon* 9: 338–70. [CrossRef]
- Bolinger, Dwight. 1968. *Aspects of Language*. New York: Harcourt, Brace and World.
- Bolkestein, A. Machtelt. 1983. Genitive and dative possessors in Latin. *Advances in Functional Grammar* 55: 91.
- Bolkestein, A. Machtelt. 2001. Possessors and experiencers in Classical Latin. *Typological Studies in Language* 47: 269–84.
- Burling, Robbins. 1992. *Patterns of Language: Structure, Variation, Change*. Leiden: Brill.
- Cacoulios, Rena Torres, and James A. Walker. 2009. The present of the English future: Grammatical variation and collocations in discourse. *Language* 85: 321–54. [CrossRef]
- Coussé, Evie, Joel Olofsson, and Peter Andersson. 2018. Grammaticalization meets construction grammar. In *Grammaticalization Meets Construction Grammar*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 1–321.
- Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. London: Pearson Education.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Croft, William, and D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Danesi, Serena, and Jóhanna Barðdal. 2018. Case marking of predicative possession in Vedic. In *Non-Canonically Case-Marked Subjects: The Reykjavík-Eyjafjallajökull Papers*. Amsterdam and Philadelphia: John Benjamins, vol. 200, pp. 181–200.
- De Smet, Hendrik, Frauke D'hoedt, Lauren Fonteyn, and Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29: 197–234. [CrossRef]
- Fedriani, Chiara. 2011. Experiential metaphors in Latin: Feelings were containers, movements and things possessed 1. *Transactions of the Philological Society* 109: 307–26. [CrossRef]
- Fillmore, Charles J. 1996. The Pragmatics of Constructions. In *Social Interaction, Social Context, and Language: Essays in Honor of Susan Ervin-Tripp*. Edited by Dan Isaac Slobin, Julie Gerhardt, Amy Kyratzis and Jiansheng Guo. Mahwah: Lawrence Erlbaum Associates, Inc., pp. 53–69.
- Fonteyn, Lauren. 2019. *Categoriality in Language Change: The Case of the English Gerund*. Oxford: Oxford University Press.
- Fried, Mirjam. 2009. Construction Grammar as a tool for diachronic analysis. *Constructions and Frames* 1: 261–90. [CrossRef]
- Gildea, Spike, and Jóhanna Barðdal. 2023. From grammaticalization to Diachronic Construction Grammar: A natural evolution of the paradigm. *Studies in Language* 47: 743–88. [CrossRef]
- Gisborne, Nikolas, and Amanda Patten. 2011. *Construction Grammar and Grammaticalization*. Oxford: Oxford University Press.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press on Demand.
- Goldberg, Adele E. 2013. Constructionist approaches. In *The Oxford Handbook of Construction Grammar*. Edited by Thomas Hoffmann and Graeme Trousdale. New York: Oxford University Press, pp. 15–31.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2015. From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics* 26: 113–47. [CrossRef]
- Hilpert, Martin. 2018. Three open questions in diachronic construction grammar. In *Grammaticalization Meets Construction Grammar*. Edited by Evie Coussé, Peter Andersson and Joel Olofsson. Amsterdam and Philadelphia: John Benjamins, pp. 21–39.
- Hoffmann, Thomas. 2013. *Abstract Phrasal and Clausal Constructions*. Oxford: Oxford Handbooks Online. [CrossRef]
- Ilioiaia, Mihaela. 2020. Productivity of the Romanian MIHI EST pattern. *Revue roumaine de linguistique* LXV: 49–67.
- Ilioiaia, Mihaela. 2021. Non-Canonical Subject Marking in Romanian: Status and evolution of the MIHI EST Construction. Ph.D. thesis, Ghent University, Ghent, Belgium.
- Ilioiaia, Mihaela. 2023. *The MIHI EST Construction: An Instance of Non-Canonical Subject Marking in Romanian*. Berlin and Boston: De Gruyter. [CrossRef]
- Ilioiaia, Mihaela, and Marleen Van Peteghem. 2021. Dative experiencers with nominal predicates in Romanian: A synchronic and diachronic study. *Folia Linguistica* 55: 255–90. [CrossRef]
- Kövecses, Zoltán. 2003. *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*. Cambridge: Cambridge University Press.
- Kurylowicz, Jerzy. 1965. The Evolution of Grammatical Categories. *Diogenes* 13: 55–71. [CrossRef]
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford: Stanford University Press, vol. 1.
- Langacker, Ronald W. 2003. Constructions in cognitive grammar. *English Linguistics* 20: 41–83. [CrossRef]
- Lauwers, Peter, and Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49: 1219–35. [CrossRef]
- Leech, Geoffrey N. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
- Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In *New Reflections on Grammaticalization*. Amsterdam and Philadelphia: John Benjamins, pp. 1–18.
- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*. Berlin: Language Science Press. First published 1982.
- Löfstedt, Bengt. 1963. Zum lateinischen possessiven Dativ. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 78: 64–83.

- Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scientia* 6: 384.
- Niculescu, Dana. 2013. The possessive dative structure. The possessive object. In *The Grammar of Romanian*. Edited by Gabriela Pană Dindelegan. Oxford: Oxford University Press, pp. 183–90.
- Noël, Dirk. 2007. Diachronic construction grammar and grammaticalization theory. *Functions of Language* 14: 177–202. [CrossRef]
- Pinkster, Harm. 2015. *Oxford Latin Syntax: Volume 1: The Simple Clause*. Oxford: Oxford University Press.
- Rostila, Jouni. 2004. Lexicalization as a Way to Grammaticalization. In *Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki, January 7–9, 2004*. Edited by Fred Karlsson. Publication No. 36. Helsinki: University of Helsinki, pp. 1–13.
- Seiler, Hansjakob. 1983. Possessivity, Subject and Object. *Studies in Language* 7: 89–117. [CrossRef]
- Stassen, Leon. 2009. Predicative Possession. In *Oxford Studies in Typology and Linguistic Theory*. Oxford: Oxford University Press.
- Stolz, Thomas, Sonja Kettler, Cornelia Stroh, and Aina Urdze. 2008. *Split Possession: An Areal-Linguistic Study of the Languages of Europe*. Studies in Language Companion Series 101; Amsterdam and Philadelphia: John Benjamins.
- Traugott, Elizabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press. [CrossRef]
- Van Peteghem, Marleen. 2017. Les structures de la douleur: Sur le marquage de l'expérienceur dans les langues romanes. In *De la passion du sens en linguistique: À Danièle Van de Velde*. Valenciennes: Presses Universitaires de Valenciennes, pp. 439–63.
- Van Peteghem, Marleen, and Mihaela Ilieoia. 2017. Nu mi-e frica de nimic: Une structure MIHI EST en roumain? In *Hommages offerts à Maria Iliescu*. Craiova: Editura Universitaria Craiova, pp. 313–27.
- Vangaever, Jasper, and Mihaela Ilieoia. 2021. Specialisation through competition: HABEO vs MIHI EST from Latin to Romanian. In *Quand le syntagme nominal prend ses marques: du prédicat à l'argument*. Reims: EPURE, pp. 243–65.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Productivity from a Metapragmatic Perspective: Measuring the Diachronic Coverage of the Low Level Lexico-Grammatical Construction Have the N (Body Part/Attitude) to \leftrightarrow <Metapragmatic Comment> Using the COHA

Chris A. Smith

CRISCO EA4255, Université de Caen, 14000 Caen, France; chris.smith@unicaen.fr

Abstract: This paper seeks to address the relation between semantics, pragmatics and the productivity of a low level lexico-grammatical construction, HAVE THE N (BODY PART/ATTITUDE) TO \leftrightarrow METAPRAGMATIC COMMENT. The question posed is how semantics affects productivity, in the generative sense of extensibility of a construction (a form meaning pairing). The method identifies the specificity and variations of the HAVE THE N (BODY PART/ATTITUDE) TO \leftrightarrow METAPRAGMATIC COMMENT construction within the pragmeme of politeness using the COHA. Hereafter, we consider how to measure the extensibility within the onomasiological frame based on the available pool of forms expressing an attitude/emotion, i.e., the coverage or attractivity of the HAVE THE N TO construction. The paper discusses the findings, namely, how to overcome methodological issues relating to a qualitative rather than quantitative approach to the constructional architecture and the relative productivity of constructions. The experimental small scale corpus study of Have the N to in the COHA suggests that a global view of constructional architecture at multiple levels should be pertinent to identifying the extensibility potential of the construction.

Keywords: grammaticalization; constructional change; productivity; coverage; diachronic; distributional semantics; COHA

Citation: Smith, Chris A.. 2023. Productivity from a Metapragmatic Perspective: Measuring the Diachronic Coverage of the Low Level Lexico-Grammatical Construction Have the N (Body Part/Attitude) to \leftrightarrow <Metapragmatic Comment> Using the COHA. *Languages* 8: 92. <https://doi.org/10.3390/languages8020092>

Academic Editors: Renata Enghels, Marlies Jansengers and Tom Bossuyt

Received: 16 January 2023
Revised: 14 March 2023
Accepted: 15 March 2023
Published: 23 March 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this paper, for the special issue “Grammaticalization across languages, levels and frameworks”, we look at constructional change rather than grammaticalization, which is, strictly speaking, “defined as the development from lexical to grammatical forms and from grammatical to even more grammatical forms” (Heine and Kuteva 2002, p. 2). “Since the development of grammatical forms is not independent of the constructions to which they belong, the study of grammaticalization is also concerned with constructions and with even larger discourse segments” (Heine and Kuteva 2002, p. 2). Thus, we take the constructional network approach (Traugott and Trousdale 2013) at its word in a maximalistic view as being a holistic (all-encompassing) apparatus following Hilpert (2013) and Goldberg (2019). As Hilpert (2013, p. 9) puts it, “[c]onstructional change is more encompassing than the changes that characterize grammaticalization”, which, arguably, is characterized by paradigmaticization. Indeed, Hilpert (2013) observes that “[c]onstructions that lack paradigmatic oppositions, that is, forms such as the way-construction, *let alone*, or *the Xer the Yer*, would in this view fall outside the scope of grammar.”

If we take a global view of the lexicogrammatical continuum and view language change as gradual and transitional, then changes involving lexis may be connected to changes involving grammatical aspects of language. If we also accept that constructions form a network of interrelated constructions of a higher and lower level as outlined in Traugott and Trousdale (2013), then lower-level constructions may have the ability to affect higher level constructions, as well as vice versa (see Hilpert 2013; Traugott and Trousdale

2013 on micro, meso and macro constructions). Given these theoretical assumptions, this paper addresses the issue of constructional change and how to determine the emergence of constructional change, as well as the productivity of such changes. Although constructional approaches are gaining ground, maybe even bolstered by this increasing popularity, the definition and properties of Cx continue to be hotly debated (see Goldberg 2019). There are also notably (and paradoxically) an increasing number of subtypes of construction grammar, such as Cx morphology or CxM (defined by Masini and Audring 2019, p. 372) as “a sign-based theory of morphology whose building blocks are constructions”), or multimodal construction grammar (Imo 2015).

Constructions¹ are defined generally as form-meaning pairings that vary in complexity and schematicity. Traugott and Trousdale (2013, p. 163) cite properties as schematicity, compositionality and productivity. Perek (2020) lists schematicity and productivity, whereas Bybee (2010, pp. 220–21) adds autonomy, prototype effects and analysability. A lexical schema is productive if it is extensible, but not necessarily open ended, as argued by Barðdal (2008, p. 1). The extension of the schema results from local analogies operating on the onomasiological (or conceptual) plane according to Traugott (2019, p. 131). Larger level constructions have been studied as being more productive (*X your way, the Xer the Yer*) which are consistent with the top-down approach usually carried out regarding the constructions.

The issues of variation and innovation (Hilpert 2013²; Goldberg 2016, 2019; Hoffmann 2020), or capacity for growth or extension of constructions, are currently being investigated from different perspectives including lexical/morphological perspectives (Audring et al. 2017; Booij and Audring 2018; Masini and Audring 2019). The objective is to avoid disregarding the importance of the lower-level lexicon which can often be overlooked or oversimplified and to account for the Cx architecture from a low level to a higher level. We believe there is a gap in the research on constructions and productivity that will benefit from a bottom-up approach to constructions, thus focusing on the lower-level lexical stratum. A bottom-up approach to constructions tends to consider the importance of low-level constructions and how they affect higher level constructions (Gyselinck 2020; Budts and Petré 2020).

The specificity and originality of our position is threefold: (1) starting with a lower-level construction belonging to the lexicogrammatical layer (Hilpert 2013, p. 202); (2) taking a qualitative approach to productivity and (3) relying on a communicative approach to constructions.

- (1) The first aspect positions this line of research firmly to use Hilpert’s (2013, p. 202) words, in the camp of “practitioners of Construction Grammar [. . .] who view low-level schemas as central to the description of linguistic knowledge” as opposed to the camp of those who “aim for abstract generalizations”. It is fundamentally a data-driven method that allows for “acknowledging the rich networks of lexical schemas and collocations [that] characterize the actual usage of grammatical constructions” (Hilpert 2013, p. 202). It should be possible to build the architecture from the bottom as well as from the top if we assume that motivational ties are multidirectional in the constructional network.
- (2) Taking things beyond frequency. It is also true that much productivity research has focused on quantitative frequency measures (Hilpert 2013), but focusing on onomasiological qualitative measures has been suggested to be a way forward (Zenner et al. 2014; Geeraerts 2016; Durkin 2016; Fernandez-Dominguez 2019; Petré 2019; Goldberg 2019; Lorenz 2020), correcting the bias in favor of more frequent abstract grammatical structures and to the detriment of less frequent lower-level lexicogrammatical structures.
- (3) Using pragmatics as an entry point for the onomasiological perspective. To test the onomasiological perspective, we follow Hilpert and Bourgeois’ (2020) main assumption that a pragmatic context motivates a set of constructions which in turn can be seen to fall into a larger structure, a metapragmatic construction so to speak. “It is a basic tenet of usage-based construction grammar (Goldberg 2006; Bybee 2010) that long-term linguistic changes originate from processes that are at work in actual com-

municative situations. [...] So far, however, relatively little work on constructional change addresses either the dialogical nature of language or the social context in which a particular construction is used P1" (Hilpert and Bourgeois 2020, pp. 97–98).

Since points (2) and (3) are interrelated, we will develop them together. As Geeraerts (2016, p. 158) argues, experiential frequency, conceptual frequency and lexical frequency should not be confused as a single effect: "The frequency of occurrence of a linguistic construct results from these three types of frequency at the same time, and talking of entrenchment as a single, unitary phenomenon obscures those differences". Petré (2019, p. 161) also argues that onomasiological space is the key to measuring entrenchment and that "the emergence of a construction is better measured by a more fine-grained quantitative analysis of the contexts in which the pre-construction material occurs".

In other words, the onomasiological approach is valuable if we want to build an accurate picture of the construction, the global architecture of constructions, i.e., the identification of constructional families (De Smet et al. (2018, p. 205). As Goldberg (2019, p. 36) puts it, "our knowledge of language forms a construct-i-con, which includes words, partially filled words (aka morphemes), and representations that are larger than single words, all represented in a complex dynamic network, much as we have long known to be true of the lexicon.) and how constructions share semantic or grammatical space." The function of interconnecting schemas and constructions in the construction is to represent the multiple motivations at work in the lexico-grammatical continuum as underlined in De Smet et al. (2018, p. 206), Audring et al. (2017) and Booij and Audring (2018).

This paper seeks to bridge the gap between the pragmatic–semantic interface and constructional productivity (see Smith 2021), and takes a maximalist lexicalist view of the construction. The pragmatic–semantic interface is operationalized using Mey's ([1993] 2001) concept of the pragme. A pragme is defined in Mey ([1993] 2001, p. 208) as an instantiated pragmatic act, or a "generalised pragmatic act" (Mey 2010, p. 2884), examples of which are "as found in invitations, bribes, co-optations, incitements, and so on—all depending on the situation through which they are defined" (Mey 2010, p. 2884). We propose to center our analysis on the pragme of insolence (or impoliteness), i.e., generalized speech acts belonging to a socio-cultural situation of face-threatening or face-preserving (see Culpeper 2011; Bousfield 2008). Our specific focus will be on the metapragmatic reaction to perceived insolence via the expression *have the N to*. The objective for this paper is to identify the specificity and variations of the Have the N to ↔ attitude construction within the pragme of insolence using the COHA or *Corpus of Historical American English*.

Our driving question is how to measure attractiveness of the construction based on the extensibility within the onomasiological frame, that is the available pool of forms expressing an attitude/emotion. The assumption behind this is that productivity should be relative to the extensibility potential of a construction. Using frequency measures exclusively tends to disregard less abstract lower level constructions, and might misrepresent their actual productivity. It is notable that Goldberg (2019) has since criticized her own reference to constructions requiring "sufficient frequency" (Goldberg 2006, p. 5) as "nonsensical", reinforcing the idea that if there is a consensus that constructions exist at multiple levels of complexity and abstractness, then it is essential that the productivity of such constructions be assessed within their own relative space. The aim is to track the emergence of the constructional architecture based on the instantiations of the construction, taking into account horizontal links (synonymy, polysemy, semantic shift) between fillers. As Wray (2017, p. 2) underlines, "[...] there is more to *usage* than just patterns of frequency. We must understand what motivates the usage, and that entails recognizing the functions that they have: both cognitive and social."

Our proposal is to carry out a lexicalist bottom-up approach to Cx starting with a low-level construction and our method will rely on a fine grained distributional semantic analysis using the COHA which we will now develop.

2. Materials and Methods

This paper will focus on the lower-level micro construction HAVE THE N TO, a construction (or collocation) of metapragmatic nature that takes a selection of nouns relating to body parts or nouns referring to emotion concepts. It is a low-level verbal lexicogrammatical structure with a single variable N consisting in a predicate structure around the verb HAVE + complement NP carrying a definite article THE and followed by infinitival structure TO. The construction falls under a general semantic abstract schema of metaphor and metonymy relating to body parts and emotions and attitudes (confidence, arrogance). The construction also belongs to the transitive ACS (argument structure construction) including an infinitival object complement.

In our conceptual space or onomasiological space perspective, we propose that this construction is a manifestation of what can be called the “pragmeme of politeness (approval/disapproval)” (see Allan 2016 on the pragmeme of insult). The pragmeme is a pragmatic routine as per Mey ([1993] 2001). Insolence (or an insolent speech act) and indignation (the reaction to a perceived insolent act) function as a pragmatic routine, i.e., a pragmeme, which in all likelihoods has a multimodal nature as insolence and indignation manifest in a broader communicative situation. Insolence is therefore more than an onomasiological field but rather a multimodal pragmeme. “The pragmeme is thus the embodied realization of all the pragmatic acts (or ‘allopracts’) that can be subsumed under it, such as the various manifestations of expressing gratitude . . . ” Mey ([1993] 2001, p. 139).

We propose that the pragmeme can be operationalized as a higher order pragmatic construction. We believe that both the level of generalization of the pragmeme and its relevance in dialogic and communicative contexts give the pragmeme its potential to be a powerful entry point into the constructional architecture. The pragmeme can be seen as a (multimodal) construction (meaning form pairing) with variables; it is a higher-level abstract construction since it can be instantiated by subconstructions. Indeed, the insolence/impoliteness pragmeme could subsume a number of constructions expressing metalinguistic reactions to perceived acts of insolence or face-threats. Within this pragmeme of insolence (or impoliteness), we can assume the existence of a whole network of interconnected subconstructions, including *have the N to*, but also *What the N, don't give me N* and *don't get Ad with me*, to name a few. Additionally, it should be noted that in terms of grammaticalization potential, insolence or the larger pragmeme which subsumes insolence or impoliteness is fundamentally intersubjective in nature (Ghesquière et al. 2015; Traugott 2015). If intersubjectivity is an orientation toward the co-speaker or the act of speech itself, markers of intersubjectivity have been found to be compatible with right periphery usage, in turn taking positions. <HAVE the N to> is a metapragmatic evaluation of the co-speakers' behavior.

In this paper, we seek to test the proportional or relative space that individual micro-constructions take in the onomasiological field of insolence or metapragmatic attitude operationalized by the pragmeme. Previous work (Smith 2021) showed that the semantic extension of body part nouns (*cheek, lip, face, nerve*) associated with insolence is related to the emergence of *have the N to*, which then generated the lexical schema Ny (*cheeky, lippy, facy, nervy*) adjective associated with the sense ‘insolent’. The working hypothesis is that these two low level constructions have emerged diachronically to associate body parts with insolence. This relative space corresponds to what Goldberg (2019, p. 51) has termed “coverage”: “the meaning and distribution of words, combinations of words, and constructions rely on the nature of our memory. In particular, partially abstracted (lossy) structured exemplars dynamically cluster within our hyper-dimensional conceptual space. [. . .] new expressions are licensed to the extent that they comfortably fit within an existing cluster.”

To test this coverage (or extensibility), this paper carries out a corpus-driven analysis of the variation of fillers in the N slot of the micro construction *Have the N to* using the diachronic balanced corpus COHA (Davies 2010) which has successfully been used for diachronic tracking studies (Hilpert 2013). The working hypothesis is the following: The

compared frequency of the lexically filled structures provides a measure of the representativeness of the exemplar within the field. From a cognitive perspective, it is supposed that the lexicon contains certain “driving” elements belonging to the core lexicon which have a higher entrenchment in the mental lexicon within a conceptual domain (Killgarriff et al.’s 2004 “dominant senses”). Although the relation between general frequency and representativeness is not straightforward, frequency measures and collocations scores can help to identify these potential drivers within the semantic space via consistency of meaning. Extensions of the pattern over time will provide evidence of the productivity of the pattern and potential evidence of semantico-pragmatic change. The method used is a corpus-driven collocational and syntactic analysis of the *have the N to* collocation (lexicogrammatical pattern) using in the COHA. Collocational behavior and lexicogrammatical patterns have been shown to be instrumental for identifying patterns of usage (Gries and Stefanowitch 2004; Perek 2020). From a diachronic perspective, tracking these behaviors over time for the period 1820 to 2010 will provide some data regarding the extensibility of the pattern, and hence evidence of the onomasiological productivity of the low-level lexico grammatical pattern. From there, we discuss how to assess the representativeness of individual fillers, using the collocate analysis to assess the proportion of usage corresponding to the insolence pragmeme.

3. Results Tracking Noun Fillers

In this section, we track the variation in noun fillers over time using the COHA to provide frequencies (overall and diachronic) and collocate profiles of specific *have the N to* constructions.

3.1. The Categories of Noun Fillers

First, we look at the variation in N fillers occupying the complement slot in the verb structure *Have the N to*. The results for the query concerning N fillers in the structure *Have the N to* in the COHA are reproduced in Table 1. The query HAVE THE N TO in capitals allowed for all variations of the lemma *have* to be included: *-had, have, has*. The following table is not exhaustive as not all forms appear; frequencies below 30 do not appear as we shall see later.

Table 1. List of N fillers for *Have the N to* in COHA and raw frequencies.

| Have the N (Concept) | | Have the N (Body Part) | |
|----------------------|--------|------------------------|--------|
| Noun | Tokens | Noun | Tokens |
| RIGHT | 3062 | NERVE | 321 |
| POWER | 1746 | HEART | 316 |
| COURAGE | 1148 | GUTS | 196 |
| OPPORTUNITY | 646 | FACE | 35 |
| HONOR | 534 | GALL | 40 |
| ABILITY | 510 | | |
| CHANCE | 483 | | |
| MISFORTUNE | 283 | | |
| TIME | 267 | | |
| STRENGTH | 253 | | |
| MONEY | 252 | | |
| AUTHORITY | 247 | | |
| CAPACITY | 188 | | |
| SENSE | 162 | | |
| AUDACITY | 161 | | |
| GOODNESS | 150 | | |
| MEANS | 141 | | |
| GRACE | 87 | | |
| RESOURCES | 87 | | |
| FORESIGHT | 73 | | |
| TEMERITY | 72 | | |

Table 1. Cont.

| Have the N (Concept) | Have the N (Body Part) |
|----------------------|------------------------|
| KEY | 71 |
| ENERGY | 65 |
| HAPPINESS | 62 |
| DECENCY | 55 |
| KINDNESS | 47 |
| IMPUDENCE | 46 |
| EFFECT | 45 |
| LUCK | 44 |
| WIT | 40 |
| PLEASURE | 36 |
| CURIOSITY | 35 |
| SATISFACTION | 34 |
| PATIENCE | 33 |
| URGE | 33 |
| WILL | 31 |

The N fillers in Table 1 are listed from the highest to lowest raw frequency in the COHA. The left column represents concept nouns (*power, opportunity, curiosity*), whereas the right column represents body part nouns.

In order to attempt a classification of types of N fillers, the concept nouns are classified conceptually in terms of the types of evaluation they describe, thus providing a pattern of possible semantic traits of the construction. We know the construction can have literal or figurative interpretations and we know the construction is associated with a metapragmatic commentary on a situation. The nature of the commentary can vary from factual to positive or negative evaluation. From the list of nouns provided, we have broadly identified five types based on a lexical semantic analysis.

- (1) Modal evaluation of right or ability: *have the right, power, authority to, capacity, strength, money, key*
- (2) Evaluation of character: *decency, kindness, heart, sense, patience, honor*
- (3) Factual: *chance, opportunity, time*
- (4) Negative evaluation of attitude: *the impudence, gall, nerve, cheek, face, audacity temerity*
- (5) Positive evaluation of attitude: *courage, foresight*

We propose that these instantiations find themselves on a politeness–impoliteness spectrum with varying degrees of judgement or evaluation (from approval to disapproval and indignation).

Whereas concept noun fillers are many, the body part fillers are far less varied. The body parts listed mostly have a metaphorical sense of impoliteness: *nerve, heart, guts, face*. Some body part nouns however are not listed (*stomach* or *cheek*) in this search but do occur, such as *cheek*, as their frequency falls below the threshold of 30; *Have the cheek* has 18 tokens, *have the stomach* 15 tokens. These frequencies are well below those of *heart, nerve, guts* which are above 30, with *nerve* and *heart* in the lead.

A further issue with the search query HAVE THE N TO is that the COHA may not reliably recognize all N slot fillers as being nouns. For instance, a search for *have the wherewithal to* produces 48 hits, which should mean these results should have been included in the previous table. However, *wherewithal* is not readily labelled as a noun, and therefore does not appear in the results. The Oxford English Dictionary (OED for hereon) labels *wherewithal* [1535] as an adverb and noun with little information regarding its formation as a compound of *where* + *withal*. The COHA occurrences show that *Have the wherewithal to* remains in use with a low frequency from 1840s to 2010s. The relative frequency shown in Figure 1 indicates that this is on the rise again between 2000–2010 after several periods of increased and decreased usage.

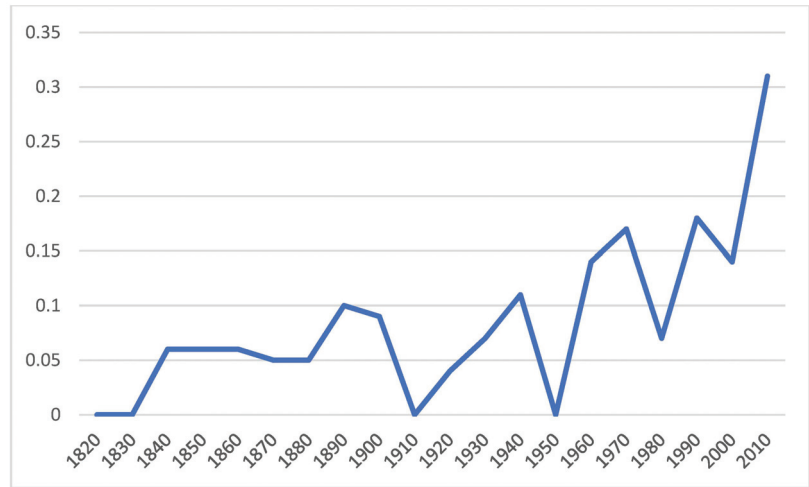


Figure 1. Relative frequency of *have the wherewithal to per decade* in COHA.

The meaning of *wherewithal* fits in with the conceptual category ability/capacity/but also means/resources/sense. The total number of occurrences of *wherewithal* average 362 with 48 uses in the pattern *have the N to*. Nearly all occurrences find *wherewithal* followed by *to* or *for*. Occasionally, the determiner *the* is omitted in some earlier occurrences as in (1). Note that the occurrences of N fillers in the corpus extract is highlighted in bold character for added visibility. The source of the COHA extract is provided as it appears in COHA (date, title and author).

- (1) “you are the real sovereigns of Castile, enjoying all the rights and revenues of royalty, while I, stripped of my patrimony, have scarcely **wherewithal** to procure the necessaries of life.” (1864. *The History of the Reign of Ferdinand and Isabella the Catholic* – Volume 1. Prescott, William Hickling)
- (2) Things I can’t understand. I could be better educated, Mr. Keele. That’s why I’ve come to you. I want some help. “I leaned back. If he found gold, he should have the **wherewithal** to get in there and back without my help” (1949. *Amazing stories: Valley of the Croen*, Tarbell, Lee)
- (3) “Hey, what’s going on here tonight?” I could not give him anything but a shrug for a reply. And try as I might, when I later kissed her goodbye at the door, I did not seem to have the **wherewithal** to feign even a little filial affection. (1974. *My life as a man*. Philip Roth)

Notwithstanding this issue of non exhaustivity of the search query, overall, as expected, since words for concepts have more variation, there are considerably more concept fillers than body part fillers (*heart, nerve, face, guts, gall*). In addition, the frequency of occurrence of the concept nouns is much higher on the whole than body part fillers, with *right, power* and *courage* occurring far more frequently than body noun fillers. The extensibility of the schema is limited since names for body parts do not evolve quickly (see Durkin 2016, who shows that borrowing tends to be restricted in certain conceptual domains).

The compared relative frequencies of highest ranking are shown in Figure 2.

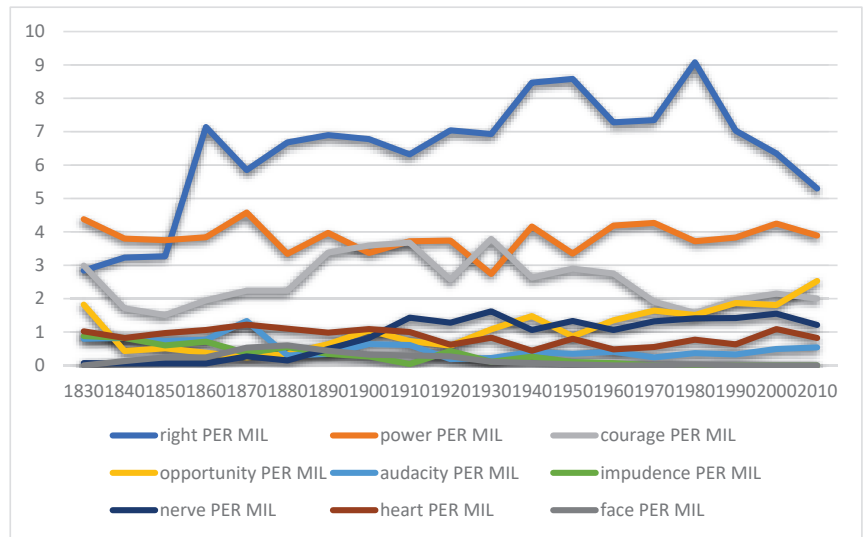


Figure 2. Relative frequencies for several *Have the N* 1820–2010 in COHA.

The relative frequencies in Figure 2 are indicative of stability over the 200-year period covered by the COHA, with some variations in the higher frequency N: *right*, *power* and *courage*. In the lesser frequency category, *nerve* is a clear winner with its increasing frequency (which we will return to). This suggests that the pattern is already well established by the 1800s.

3.2. Metapragmatic Comment Fillers

Based on the variation in concept fillers shown previously in Table 1, it also appears that a proportion fits the metapragmatic comment on the courage/insolence or approval/disapproval spectrum: namely *courage*, *strength* (on the positive spectrum), *audacity* and *temerity* (on the negative spectrum). Table 2 lists the frequency of some of these N fillers.

Table 2. Have the N (+/– impudence/courage) in COHA.

| Have the N | Tokens |
|------------------|--------|
| <i>Courage</i> | 1148 |
| <i>Strength</i> | 253 |
| <i>Audacity</i> | 161 |
| <i>Temerity</i> | 72 |
| <i>Impudence</i> | 46 |

These fillers are comparable semantically to the far less frequent body part fillers shown in Table 3 carrying metonymical meanings for insolence, impudence, impoliteness and courage.

Table 3. Have the N (body part) in COHA.

| Have the N | Tokens |
|--------------|--------|
| <i>Nerve</i> | 321 |
| <i>Heart</i> | 316 |
| <i>Guts</i> | 196 |
| <i>Face</i> | 25 |
| <i>Gall</i> | 40 |

The development of the politeness/impoliteness association of body parts historically appears related to the emergence of construction <Have the N to> insolence as proposed in Smith (2021). If we posit a micro schema 1: Have the N to insolence and micro-schema 2 Ny adj ↔ insolence, the emergence of schema 2 is based on the inheritance of information relating to Schema 1. Table 4 presents the attestation dates of Ny (schema 2) compared N (schema 1) based on OED data.

Table 4. Attestation dates of N and N-y with the sense ‘insolent’ in the OED (Smith 2021).

| Body Part Name | Other Figurative Senses | Schema 1 | Schema 2 |
|----------------|-------------------------|---|-------------------------------|
| Lip OE | Yes | Give lip 1821 | Lippy 1875 colloq or dial |
| Cheek OE | No | Have the cheek 1823 Give cheek 1825 | Cheeky 1838 impudent |
| Nerve OE | Yes | Have the nerve 1887 | Nervy 1896 US colloq |
| Face 1300 | Yes | Have the face 1562 | Facy 1607 obs except dial |
| Front 1290 | Yes | Have the front 1653 | *fronty not attested |
| Brow OE | Yes | Have brow 1642 (obs) | *browy not attested |
| Chin | Yes | Have chin 1877 Give chin 1877 | Chimmy 1883 (sense talkative) |
| Mouth OE | Yes | Be a mouth 1699 Be all mouth (insolent) | Mouthy 1589 (polysemic) |
| Forehead OE | Yes | Give mouth 1825 (express) Have the forehead 1564 (obs) | *foreheady not attested |

Diachronic lexicographic data from the OED support this hypothesis that <Have the N to> construction may have led to the emergence of the morphological lexical construction, Ny adjective—insolence (*cheeky, nervy, gutsy, *facy*) from the 1830s. *Have the cheek, have the nerve, have the guts* and *have the face* all predate the attestation of denominal adjectives *cheeky, nervy, gutsy* and **facy* (which has ultimately become obsolete).

Of course, the metaphorical or metonymical shift involved in the extension of the meaning of body part names to insolence no doubt has a deeper cognitive origin, since body parts and facial expressions are viewed as the manifestation (or origin) of interpersonal emotions. Body part names are well known to be particularly susceptible to semantic shifts (for instance, see Niemeier 2000; Hilpert 2006).

The existence of a number of competing *Have the N* expressions (both concept, emotion and body part nouns) shows that extensions are not blocked by synonymy relations (on the contrary, similarity is a factor in extension as argued in Goldberg 2019). Indeed, despite the frequency variations, some of these less frequent expressions may be ambivalent on the evaluation of the politeness–impoliteness spectrum. The ambivalence hinges on the indignation/disapproval versus approval rating, relating to the value attributed by the speaker to another person’s actions or behavior. In (4), *nerve* takes a negative disapproval value, compared to (5) where *nerve* is associated with courage rather than insolence:

- (4) Ah, Mrs. Pampinelli. Mrs. Pampinelli. Look at her. All over the front page again. You know, she and I are really in the same business. I take pigs and turn ‘em into sausages... and she takes our citizens and turns ‘em into hams. Her last show was so bad . . . I didn’t think she’d have the **nerve** to try and put on another one. Oh, she takes a lot of our young folks . . . and she turns their head by telling them they’d be great on the stage. (1935 *Doubting Thomas*, TV Movie)
- (5) That has been deliberate on my part. I better say this now or I won’t have the **nerve** to say it later. (2001. *Sun Rising On the Hill District*, Robert Penny)

Although *guts* in (6) is interpreted as courage, most occurrences are compatible with a metapragmatic intention of daring or goading from the speaker as in (7). It is notable that many occurrences of *have the guts* to feature negative polarity as in (7) and (8).

- (6) But enlarged pictures of the individuals, separated from the total, disporting themselves in lewd, naked positions would do the job. Clearly the police must put a stop

to this. He would have every organization in the universe dedicated to dictating the morals of others on his side. No politician would have the **guts** to stand up in opposition. (1960 *Eight Keys to Eden*, Mark Clifton)

- (7) And you're not. Let's face it, kid. You don't have the **guts** to kill me. (1999. *Mystery Men*, TV Movie)

Despite the apparent synonymy of the patterns, these corpus extracts suggest that metapragmatic differences are present from one filler to the next. Behind their uses lie the question whether the ambivalence insolence/courage (disapproval/approval) is open, i.e., a matter of the content of the infinitival clause, or whether some fillers have a preferential negative or positive reading. Do *Have the N body part* constructions have a higher ambivalence reading than the *Have The N concept* constructions? In order to examine the question, we now turn to a collocate analysis in the COHA.

3.3. Collocate Patterns in the Infinitival Clause

A search for collocate patterns +1 to +5 is carried out to determine what types of verbs are significantly found in the *Have the N to* complement clause. Reducing the search to collocates in the +1 to +5 position does not target only verbs but all collocates found in this segment. The collocate search was carried out for specific *have the N to* structures for comparison: *heart, nerve, stomach, cheek, courage, temerity, audacity, impudence*. The results show a high frequency of collocates corresponding to speech act verbs and position taking verbs, such as *say, tell, speak, ask, admit, call, laugh* and *disappoint*.

The *raw* frequencies of collocates are compared to MI scores (Mutual Information) which provide a measure of the statistical likelihood of co-occurrence of the collocates with the target word (or its "proximity" see Goldberg 2019). This score therefore mitigates the frequency to provide some indication of the probability of co-occurrence. We will now compare collocate rankings based on a decreasing MI score. We should keep in mind that the collocates were not filtered to verbs only, although a large majority of hits correspond to verbs. We will not be discussing the noun collocates, many of which are pronouns (*me/themselves*) but have left the data in the tables for transparency.

We will proceed by analyzing several sets of synonymous patterns belonging to the insolence pragmeme; we begin with the body part filler *heart* versus the concept non filler *courage* representing positive evaluation (praise). We then compare three synonymous concept noun fillers (*impudence, audacity, temerity*), and finally, we compare two body part fillers representing negative evaluation (*gall* and *nerve*).

3.3.1. Collocates of Have the Heart and Courage

We turn to *have the heart* (1820–2010) and *have the courage* (1820–2010) which appear to be somewhat synonymous based on lexicographic information (OED); *Have the courage to* has 1194 tokens, *have the heart to* has 400 tokens. Note that this is slightly more than the figures provided in Table 1, but some variations were not taken into account in Table 1 due to thresholds for the general search for HAVE THE N TO.

Table 5 shows that top ranking MI scores correspond to more infrequent occurrences such as *reprimand, disappoint* and *scold*, whereas the more frequent collocates *ask* and *tell* result in much lower significance scores. This is expected due to the high overall frequency of general verbs such as *tell* and *ask*.

Similarly, Table 6 provides the list of collocates of *Have the courage* in a window of +1 to +1 to the right of the construction. These collocates are once again classified based on the MI score of the collocation.

The collocates of the pattern *have the courage* (which is almost three times more frequent than *have the heart*) follow a similar pattern. Frequent collocates such as the general verbs *tell, ask* and *speak* are not the highest ranking in terms of MI score. Instead, the more specific verbs *defy, confront* and *resist* find themselves at the top of the significance list, suggesting *have the courage* is followed by a verb indicative of disagreement/opposition (*resist, refuse, confront*) or risk taking (*risk, try*).

Table 5. Collocates of *Have the heart* to +1 to +5 in COHA.

| Collocates of <i>Have the Heart</i> | Frequency | Relative Freq | Mi Score |
|-------------------------------------|-----------|---------------|----------|
| DISAPPOINT | 6 | 0.35 | 10.01 |
| REPRIMAND | 2 | 0.31 | 9.84 |
| SCOLD | 3 | 0.26 | 9.6 |
| ASSAIL | 2 | 0.24 | 9.47 |
| MEDITATED | 2 | 0.18 | 9.07 |
| DISTURB | 7 | 0.12 | 8.5 |
| REFUSE | 10 | 0.07 | 7.7 |
| SPOIL | 3 | 0.06 | 7.53 |
| REPROACH | 2 | 0.05 | 7.12 |
| COMMIT | 3 | 0.04 | 6.94 |
| UNDERTAKE | 2 | 0.04 | 6.81 |
| WAKE | 6 | 0.03 | 6.58 |
| DENY | 4 | 0.03 | 6.43 |
| ROB | 2 | 0.03 | 6.24 |
| TREAT | 3 | 0.02 | 5.88 |
| SHOOT | 4 | 0.02 | 5.78 |
| HARM | 3 | 0.02 | 5.76 |
| TELL | 52 | 0.02 | 5.74 |
| THROW | 5 | 0.02 | 5.62 |
| SUGGEST | 3 | 0.02 | 5.6 |
| DESERT | 3 | 0.02 | 5.56 |
| SUSPECT | 2 | 0.02 | 5.51 |
| WRITE | 6 | 0.01 | 5.26 |
| KILL | 6 | 0.01 | 5.14 |
| TURN | 12 | 0.01 | 5.09 |
| BLOW | 3 | 0.01 | 4.98 |
| BLAME | 2 | 0.01 | 4.96 |
| STRIKE | 3 | 0.01 | 4.84 |
| MENTION | 2 | 0.01 | 4.69 |
| ENJOY | 2 | 0.01 | 4.61 |
| BREAK | 4 | 0.01 | 4.57 |
| ASK | 8 | 0.01 | 4.51 |
| PICK | 2 | 0.01 | 4.17 |
| LAUGH | 2 | 0.01 | 4.02 |
| ENTER | 2 | 0.01 | 3.95 |
| DRIVE | 2 | 0 | 3.74 |
| SAY | 17 | 0 | 3.62 |
| EAT | 2 | 0 | 3.6 |
| SEND | 2 | 0 | 3.53 |
| STAND | 3 | 0 | 3.37 |
| LEAVE | 4 | 0 | 3.27 |
| STOP | 3 | 0 | 3.27 |
| AWAY | 10 | 0 | 3.2 |
| HIM | 47 | 0 | 3.17 |
| POOR | 3 | 0 | 3.05 |

As expected, based on the OED data for the noun *courage*, *have the courage* has a consistently positive value assessment as in (8) and (9) extracts, although this is mitigated by the larger (often dialogic) context:

- (8) We want to know if you have the **courage** to print in your newspaper exactly what's wrong with San Francisco. That's right. I've waited a long time for this. (1935. *Barbary Coast*, TV Movie)
- (9) What is it? I'm disobeying my own orders. You remember how I'd listen to the broadcasts? About all those people who've escaped? About that railroad train that broke through? I knew that was what you had in your mind. I've known it for a long time. But I never dreamed you'd ever have the **courage** to do it. (1953. *Man on a Tightrope*, TV Movie.)

Table 6. Collocates of *have the courage* to +1 to +5 in COHA.

| Collocates of <i>Have the Courage</i> | Frequency | Rel. Frequency | MI Score |
|---------------------------------------|-----------|----------------|----------|
| DEFY | 8 | 0.31 | 8.29 |
| CONFRONT | 5 | 0.18 | 7.46 |
| RESIST | 9 | 0.08 | 6.39 |
| MISTAKES | 5 | 0.07 | 6.09 |
| CONFESS | 7 | 0.07 | 6.03 |
| REFUSE | 9 | 0.06 | 5.97 |
| ADMIT | 12 | 0.05 | 5.59 |
| SPEAK | 26 | 0.03 | 4.88 |
| ATTEMPT | 11 | 0.03 | 4.82 |
| ASK | 28 | 0.03 | 4.74 |
| STAND | 23 | 0.03 | 4.73 |
| OFFER | 9 | 0.02 | 4.52 |
| TELL | 58 | 0.02 | 4.32 |
| TRY | 18 | 0.02 | 4.24 |
| RISK | 5 | 0.02 | 4.24 |
| FIGHT | 11 | 0.02 | 4.23 |
| ENTER | 7 | 0.02 | 4.18 |
| KILL | 9 | 0.02 | 4.15 |
| CARRY | 8 | 0.02 | 4.1 |
| WRITE | 7 | 0.02 | 3.91 |
| ACT | 12 | 0.01 | 3.79 |
| THEMSELVES | 16 | 0.01 | 3.77 |
| FACE | 34 | 0.01 | 3.68 |
| DIE | 7 | 0.01 | 3.65 |
| SAY | 48 | 0.01 | 3.54 |
| LIVE | 12 | 0.01 | 3.54 |
| WALK | 7 | 0.01 | 3.53 |
| TRUST | 5 | 0.01 | 3.48 |
| REACH | 5 | 0.01 | 3.43 |
| TAKE | 35 | 0.01 | 3.19 |
| LEAVE | 11 | 0.01 | 3.15 |

3.3.2. Collocates of *Have the Impudence, Audacity and Temerity*

Based on the OED data for those entries, the French loan nouns *impudence*, *audacity* and *temerity* appear to be more compatible with a negative reading. The collocates of *have the impudence* to are shown in Table 7:

Table 7. Collocates of *have the impudence* to +1 to +5 in COHA.

| Collocates of <i>Have the Impudence</i> | Frequency | Rel Frequency | MI Score |
|---|-----------|---------------|----------|
| WINK | 2 | 0.07 | 9.48 |
| REFER | 3 | 0.04 | 8.59 |
| GLARE | 2 | 0.03 | 8.5 |
| INSULT | 2 | 0.03 | 8.24 |
| MAINTAIN | 2 | 0.01 | 6.88 |
| CLAIM | 2 | 0.01 | 6.17 |
| ASK | 7 | 0.01 | 6.13 |
| SPEAK | 4 | 0 | 5.56 |
| CALL | 6 | 0 | 5.35 |
| SEND | 2 | 0 | 5.33 |
| PRESENT | 3 | 0 | 4.57 |
| TRY | 2 | 0 | 4.45 |
| TELL | 6 | 0 | 4.43 |
| ME | 24 | 0 | 4.1 |
| SAY | 6 | 0 | 3.93 |
| SET | 2 | 0 | 3.68 |
| FACE | 3 | 0 | 3.57 |

In the case of *have the impudence* (1820–1960), a similar pattern to *have the courage/heart* emerges with verbs of speech such as *tell*, *call* and *speak*, having higher frequencies and lower MI scores. At the top of the MI ranking are *wink*, *refer*, *glare* and *insult*, more specific verbs implying some risk-taking and boundary-pushing on the part of the co-speaker.

Corpus examples for *impudence* show a high level of heightened negative assessment in (10) and (11), which is the most recent occurrence in the COHA.

- (10) “Miss!” roared the old man, bringing down his cane with a resounding thump upon the floor; “miss! how dare you have the impudence to face me, much less the – the – the **assurance!** – the **effrontery!** – the **audacity!** – the **brass!** to speak to me!” (1867. *Hidden Hand*. Southworth, Emma.)
- (11) I am surprised that after your insolent references to myself, Sir Osbert and Mr. Sacheverell Sitwell her younger brother made in verse some years ago, you should have the **impudence** to invite me to waste my time at your show. (*Time Magazine*: 1961/10/12.)

Tables 8 and 9 representing *have the temerity* (1830–2010) and *have the audacity* (dates) confirm the larger pattern of collocates representing conflict and disagreement. (12) shows the righteous indignation implied in the expression, which appears to be being phased out based on diachronic occurrences in the COHA.

Table 8. Collocates of *have the temerity* to +1 to+5 in COHA.

| Collocates of <i>Have the Temerity</i> | Frequency | Rel Frequency | MI Score |
|--|-----------|---------------|----------|
| DISAGREE | 2 | 0.07 | 9.13 |
| EXECUTE | 2 | 0.06 | 8.83 |
| SUGGEST | 9 | 0.05 | 8.51 |
| CHALLENGE | 4 | 0.03 | 7.66 |
| DECLARE | 2 | 0.02 | 7.31 |
| VENTURE | 2 | 0.02 | 7.24 |
| MAINTAIN | 3 | 0.02 | 6.99 |
| ATTEMPT | 6 | 0.02 | 6.85 |
| DENY | 2 | 0.01 | 6.75 |
| APPROACH | 3 | 0.01 | 6.3 |
| EXPRESS | 2 | 0.01 | 6.1 |
| ADDRESS | 2 | 0.01 | 5.85 |
| WRITE | 3 | 0.01 | 5.59 |
| DIRECTLY | 2 | 0.01 | 5.46 |
| ASK | 6 | 0.01 | 5.42 |
| LAUGH | 2 | 0.01 | 5.34 |
| ENTER | 2 | 0.01 | 5.28 |
| QUESTION | 3 | 0 | 4.21 |

- (12) And you mean to tell me the Germans have the **temerity** to attempt a raid in the very mouth of the Thames?” Lord Hastings nodded. “They certainly have,” he said quietly. (1915. *The Boy Allies Under the Sea*, Robert Drake.)

(13) is the most recent occurrence in the COHA, and illustrates the strong domineering judgment implied by this more outdated expression.

- (13) How dare those state workers take it in the shorts for a full career of service to the public and then have the **temerity** to expect a decent pension? J. Brandeis Sperandeo, Denver (2010. *Denver Post Open Forum*)

Table 9. Collocates of *have the audacity* to +1 to +5 in COHA.

| Collocates of <i>Have the Audacity</i> | Frequency | Rel Frequency | MI Score |
|--|-----------|---------------|----------|
| INSULT | 4 | 0.06 | 8.22 |
| PROPOSE | 4 | 0.05 | 7.91 |
| ASSERT | 2 | 0.04 | 7.68 |
| INVITE | 2 | 0.03 | 7.24 |
| ADOPT | 2 | 0.03 | 7.05 |
| DEMAND | 7 | 0.02 | 6.8 |
| LAUGH | 7 | 0.02 | 6.61 |
| PAYMENT | 2 | 0.02 | 6.52 |
| DENY | 2 | 0.01 | 6.22 |
| CLAIM | 3 | 0.01 | 5.74 |
| ATTEMPT | 4 | 0.01 | 5.73 |
| ADDRESS | 2 | 0.01 | 5.31 |
| RAISE | 2 | 0.01 | 5.31 |
| CITIZENS | 2 | 0.01 | 5.19 |
| ASK | 7 | 0.01 | 5.11 |
| STRIKE | 2 | 0.01 | 5.05 |
| ESCAPE | 2 | 0.01 | 4.95 |
| SIT | 3 | 0 | 4.66 |
| EXPECT | 2 | 0 | 4.62 |
| DRIVE | 2 | 0 | 4.53 |
| CALL | 6 | 0 | 4.33 |
| SAY | 16 | 0 | 4.32 |
| TELL | 10 | 0 | 4.15 |
| TRY | 3 | 0 | 4.02 |
| HOLD | 3 | 0 | 3.98 |
| STAND | 2 | 0 | 3.57 |
| CHURCH | 2 | 0 | 3.55 |
| BRING | 2 | 0 | 3.44 |
| THINK | 9 | 0 | 3.4 |
| FIRE | 2 | 0 | 3.33 |
| TURN | 2 | 0 | 3.29 |
| RUN | 2 | 0 | 3.29 |
| USE | 3 | 0 | 3.21 |

Have the audacity (1830–2010) is still in use, with a stronger pattern of usage until 2010 than *have the temerity* which appears to be phased out, becoming archaic. In (15), the extract taken from another forum is also suggestive of sarcasm or irony (see Garmendia 2018; Lehmann 2023), defined as an incongruence between content and opinion.

- (14) But a true word, fresh from the lips of a true man, is worth paying for, at the rate of eight dollars a day, or even of fifty dollars a lecture. The taunt must be an outbreak of jealousy against the renowned authors who have the **audacity** to be also orators. (1859. *Autocrat of the Breakfast Table*, Oliver Holme)
- (15) don't get why they are so incensed that American citizens would have the **audacity** to want to know the truth about the war in Iraq. It is our right in a democracy to question and demand the truth from our leaders. It is dangerous not to question and demand. (2005. *San Francisco Chronicle: Letters to the editor.*)

3.3.3. Collocates of *Have the Gall and Nerve*

Have the gall (1910–2010) and *Have the nerve* (1830–2010) appear semantically similar; however, the ambivalence of *nerve* is not present with *gall*, which appears lexically far more on a derogatory scale. The collocates of *gall* in Table 10 however do not clearly show negative polarity, with a similar collocate list to the other patterns.

Table 10. Collocates of *have the gall* +1 to +5 in COHA.

| Collocates of <i>Have the Gall</i> | Frequency | Rel Frequency | MI Score |
|------------------------------------|-----------|---------------|----------|
| ME | 11 | 0 | 3.09 |
| SAY | 7 | 0 | 4.27 |
| COME | 7 | 0 | 3.97 |
| TELL | 6 | 0 | 4.55 |
| ACTUALLY | 3 | 0.01 | 5.86 |
| ASK | 3 | 0 | 5.02 |
| CALL | 3 | 0 | 4.47 |
| ACCUSE | 2 | 0.08 | 9.83 |
| CLAIM | 2 | 0.01 | 6.29 |
| THROW | 2 | 0.01 | 6.23 |
| WALK | 2 | 0 | 5.23 |
| STAND | 2 | 0 | 4.71 |
| SPEAK | 2 | 0 | 4.68 |
| TRY | 2 | 0 | 4.57 |
| SELF-DEFENSE. | 1 | 4.35 | 15.59 |
| HELIOGRAPH | 1 | 3.85 | 15.41 |
| DISENFRANCHISE | 1 | 3.7 | 15.35 |
| RISIBLE | 1 | 1.89 | 14.38 |
| OBJETS | 1 | 0.81 | 13.17 |
| HUFFY | 1 | 0.63 | 12.8 |
| SAUNTER | 1 | 0.33 | 11.85 |
| POTTERS | 1 | 0.25 | 11.48 |
| CONFISCATE | 1 | 0.25 | 11.47 |
| PLATITUDES | 1 | 0.2 | 11.12 |
| UNINVITED | 1 | 0.19 | 11.05 |
| SMUGGLE | 1 | 0.18 | 10.95 |
| DEPLORE | 1 | 0.11 | 10.28 |

The occurrence in 16 shows high levels of animosity in the form of a rhetorical question:

- (16) The other flushed a deep red. “Until then,” said Barney easily, “I shall be forced to regard you and your officers as prisoners of war. “Johns laid his hand on his sword at this Yankee impudence. “You have the **gall** to let the matter go unreported all day? And to hold me aboard?” (1951. *Captain Barney: A novel*. Jan Westcott)
- (17) When Susan Weiner first ran for mayor here five years ago, at least one local was appalled that a New York Republican newcomer would have the **gall** to think she could just waltz in and take over this genteel old Southern city. (1995. *Atlanta Journal Constitution Our Southern Yankee*, Jingle Davis)

(16) and (17) show that the type of verb in the infinitival complement slot need not be particularly damning for a derogatory interpretation to emerge. The triggers of indignation lie in the metapragmatic context between the speakers dictated by their socio-cultural positions and functions.

Have the nerve (1830–2010) on the other hand is both more frequent and less polarizing than *have the gall*. The collocates ranked according to their MI score in Table 11 show, however, that the construction remains somewhat predictable.

Table 11. Collocates of *Have the nerve* +1 to +5 in COHA.

| Collocates of <i>Have the Nerve</i> | Frequency | Rel Frequency | MI Score |
|-------------------------------------|-----------|---------------|----------|
| HOUSEWIVES | 2 | 0.14 | 8.49 |
| CONTRADICT | 2 | 0.14 | 8.47 |
| TACKLE | 3 | 0.09 | 7.74 |
| UNDERTAKE | 2 | 0.04 | 6.56 |
| ASK | 36 | 0.03 | 6.44 |
| INVITE | 2 | 0.03 | 6.21 |
| JUMP | 4 | 0.03 | 6.14 |
| DRAG | 2 | 0.03 | 5.98 |
| PROPOSITION | 2 | 0.02 | 5.6 |
| EXAMINE | 2 | 0.02 | 5.57 |
| RESIST | 2 | 0.02 | 5.55 |
| TRY | 17 | 0.02 | 5.49 |
| BOTHER | 2 | 0.02 | 5.36 |
| REFUSE | 2 | 0.01 | 5.13 |
| SHOOT | 3 | 0.01 | 5.12 |
| PUSH | 2 | 0.01 | 4.94 |
| SUGGEST | 2 | 0.01 | 4.77 |
| WRITE | 5 | 0.01 | 4.75 |
| CALL | 16 | 0.01 | 4.72 |
| STAND | 9 | 0.01 | 4.71 |
| WEAR | 3 | 0.01 | 4.68 |
| TELL | 29 | 0.01 | 4.65 |
| THROW | 3 | 0.01 | 4.64 |
| KISS | 2 | 0.01 | 4.49 |
| STICK | 2 | 0.01 | 4.44 |
| CHARGE | 4 | 0.01 | 4.41 |
| CLAIM | 2 | 0.01 | 4.12 |
| BRING | 6 | 0.01 | 3.99 |
| RIDE | 2 | 0.01 | 3.99 |
| PICK | 2 | 0.01 | 3.93 |
| KILL | 3 | 0.01 | 3.9 |
| ANYONE | 3 | 0.01 | 3.87 |
| START | 4 | 0.01 | 3.79 |
| LAUGH | 2 | 0.01 | 3.77 |
| DIE | 3 | 0.01 | 3.76 |
| ATTEMPT | 2 | 0.01 | 3.7 |
| WALK | 3 | 0 | 3.64 |
| SAY | 20 | 0 | 3.61 |
| PUT | 12 | 0 | 3.55 |
| ACT | 4 | 0 | 3.54 |
| GO | 24 | 0 | 3.47 |
| CARRY | 2 | 0 | 3.43 |
| BUY | 2 | 0 | 3.38 |
| RUN | 4 | 0 | 3.25 |
| COME | 17 | 0 | 3.08 |
| ME | 49 | 0 | 3.07 |

In the case of *have the nerve*, the collocate *ask* is ranked higher on the MI score than other verbs of speech, *tell*, *claim* and *call*, indicating a higher attractivity for the slot than in the case of *have the N* (concept). The occurrences of *ask* show the ambivalence of *have the nerve*. Compare the positive evaluation in (18) with (19) where *have the nerve* is indicative of disbelief and negative evaluation:

- (18) Of course I had to go, after that – and I nearly killed myself. I thought I was pretty good to even try it. Nobody else in the party tried it. Well, afterward Rosalind had the **nerve** to ask me why I stooped over when I dove.’ It didn’t make it any easier,’ she said,’ it just took all the courage out of it. (1920. *This Side of Paradise*. Scott Fitzgerald.)

- (19) That does it, I've had it." Harry stilled. "What's wrong?" "You have the **nerve** to ask me what's wrong? After telling me that you're not going to approve Duncan's proposal?" (1997. *Absolutely, positively*, Krentz, Jayne Ann.)

The other verbs in the infinitival position that are ranked higher are *contradict* and *tackle*, verbs of speech acts indicative of conflict or disagreement. In the case of (20), they are indicative of a sense of entitlement or high regard for one's ability (self-assurance).

- (20) Skeptics who include other law enforcers have had the **nerve** to contradict Hoover on the point, and to suggest that the crimes record division of the FBI is also its public relations operation. (1970 *The New Republic*: 11/28/70.)
- (21) But later on I may tip you off to a lot of things that Morrow did that a Diplomat didn't do. In fact, he told me one time that he didn't believe he would have had the **nerve** to tackle the job, if it was not that he had the example of Diplomats to watch, so that he could do the opposite and feel sure of success. (1932. *Saturday Evening Post: Letters of a Self-Made Diplomat to Senator Borah*)

The success of the *have the nerve* pattern historically may be accounted for by the ambivalence in its usage, compared to other patterns. We now take a closer look at body part fillers and their usage profiles in the COHA.

3.4. Diachronic Variation in Fillers: Density

3.4.1. Concept Noun Fillers over Time

As we have observed, concept noun fillers are far more frequent than body part fillers. This is expected given the wide range of concept nouns available, and the much more limited availability of body part nouns. The frequency per mil of *Have the N to* appears to be relatively stable over the COHA period as shown in Figure 3.

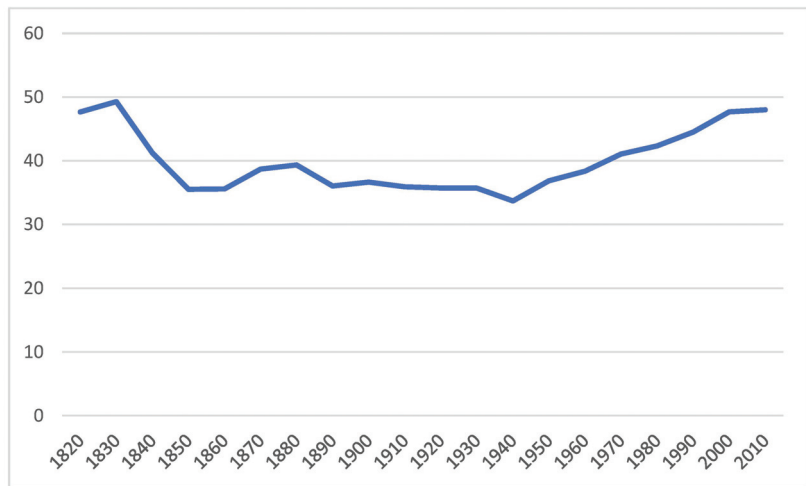


Figure 3. Relative frequency per decade of HAVE THE N TO in COHA (1820–2010).

This means the pattern itself is likely established by 1810 based on concept nouns of attitude and emotion, with the development of body part fillers from the 1800s. This can be verified based on OED attestations for the senses of body parts as shown in Smith (2021). However, we can question how the usage of filler nouns in the COHA has evolved over the course of the 200 year span of the corpus.

As is well known, one of the main issues in tracking change is establishing if variation can be interpreted as shift, and if some directionality can be established, for instance towards higher intersubjective readings. In the case of *have the N to* patterns, collocates appear remarkably similar with speech act verbs being prevalent. What distinguishes the

patterns are specific metapragmatic readings and the level of ambivalence of readings. More than frequency itself, we can argue it is the prototypicality of the association of a pattern and a reading that indicates the strength of the pattern, therefore, its onomasiological productivity.

The capacity for innovation of *have the N to* is inherently limited to the class of filler nouns of concepts and body parts relating to politeness and impoliteness but within that, field variation is present and extensibility is possible although not broad.

3.4.2. Body Part Fillers over Time

In this subsection, we look at diachronic changes in frequency starting with body part nouns. A closer look at the frequencies and dates of attestation in the COHA for *have the N (body part)* reveals higher frequencies than found in Section 3.1 based on the queries *Have the N to*. There is no specific search query enabling us to select body part nouns in the complement position N. A manual search for each fully instantiated lemmatized expression *Have the N to* provided seven expressions: *Have the nerve to* (473), *have the guts to* (305), *have the gall to* (105), *have the face* (71), *have the cheek to* (18), *have the stomach to* (15), *have the balls to* (58). These results shown in Table 12 include all verbal forms of *have* (*having*, *'ve*, *had*, etc.).

Table 12. Compared frequency of body part fillers *have the N to*.

| Lemmatized Expression | Tokens |
|----------------------------|--------|
| <i>have the guts to</i> | 305 |
| <i>have the stomach to</i> | 15 |
| <i>have the nerve to</i> | 473 |
| <i>have the gall to</i> | 105 |
| <i>have the cheek to</i> | 18 |
| <i>have the face to</i> | 71 |
| <i>have the balls to</i> | 58 |

Figure 4 shows the diachronic distribution of the tokens for each of the seven patterns in the COHA, providing a timeline of their compared usage in the corpus over time.

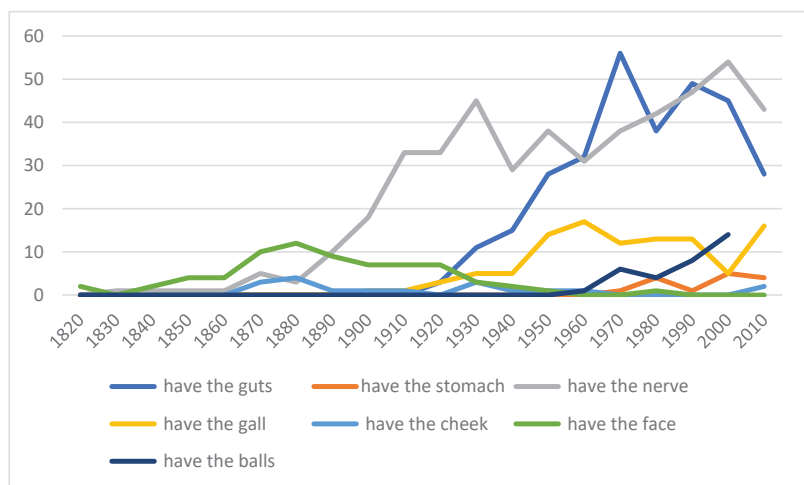


Figure 4. Frequency evolution of seven *Have the N (body part) to* patterns in COHA.

Of the seven patterns, *have the face* emerged earliest and has died out from the 1940s. *Have the cheek* is attested earlier than *have the stomach*, which appears to be a recent extension (1970). *Have the nerve* has the strongest progression from the 1880s, and *have the guts* has increased steadily from the 1930s until the present. *Have the stomach* emerges latest in the

1970s, with *have the balls* emerging around the same period. The first occurrence of *have the balls* [1969] is situated in a sexually explicit fictional dialogue in (22).

- (22) Nobody will see you, Darrell says. It's too cold for doctors or nurses to be out and nobody else cares. And if you actually do have the **balls** to fuck her you get to be leader of our club for as long as you're here, Corky says as the others solemnly nod their approval. That's true, Timmy, Darrell says. That's how Big Jim got to be boss and that's how I got to be second-in-command. They really do take me for an idiot, I think to myself. (1969. *Listen to the Silence*, David W. Elliot)

Most of the subsequent occurrences are from TV or movie sources exhibiting a high level of oral colloquialism and tolerance for vulgar language (with other references to dysphemistic abuse terms).

- (23) Oh, really? "Marcus the Magnificent." The boy, the genius, the legend. That's their name, okay? Not mine. I'm willing to bet the only reason you came to Burrows in the first place, and not one of those Ivy League schools, is because you didn't have the **balls** to measure yourself against real competition. You're a phony. A fake. Thank you for your time, Professor, and your, uh, constructive criticism. (2018. *The Truth About the Harry Quebert Affair*, TV Movie)

At the opposite end of the scale, *have the cheek* appears to have the lowest altogether frequency of usage despite a longer lifespan (1870–). (24) illustrates one of the earliest occurrences, and (25) one of the later occurrences which have become scarce in the COHA.

- (24) With all the channels of truthful information thus open and unobstructed, you preferred to get what you wanted from a spy. Mr. Howard has the **cheek** to proclaim that during the "labors" of his committee, instead of acting upon honest and legitimate evidence, he sent inquiries to this secret informer, who answered by giving in formation of "great importance," but his communications "were always indirect and anonymous!" (1871. *Galaxy* February 1871: 257–276; *Mr Black to Mr Wilson*. J.S. Black.)
- (25) While he was getting undressed last night she'd had the **cheek** to say, "Those pants are getting tired". (2011. *Night Thoughts*. Helen Simpson)

From this, it could be suggested that *have the cheek* is on a path of regression being outperformed by more recent and more dysphemistic terms such as *stomach* and *balls*. However, the issue of the variety of English is certainly in question, as is the type of data provided by a specific corpus. A search for *have the cheek* in the contemporary OEC corpus or Oxford English Corpus (2000–2010), via Sketch Engine, shows *have the cheek* to have a non-negligible relative frequency of 0.04 per mil. (26) is taken from a rather scathing literary review with a tone of strong condescension and disapproval:

- (26) With the exception of Amy Sohn and Benjamin Anastas (the author of the cult novel, *An Underachiever's Diary*) and his terrific essay here, "An Unexamined Life", most of the writers in this anthology have the **cheek** to pronounce Salinger's prose variously as "failed poetry", "brilliant writing steam rolling everything", "original without being good", "workmanlike", relentlessly middleclass and middlebrow—while their own prose is artless, ungenerous, mediocre. (May 2002. *The Hindu: Literary Review*.)

3.4.3. Issues with Low Frequency

Low frequencies raise a number of methodological issues in corpus linguistics. In addition to the issue of the data pool, raw frequency patterns can also be deceptive in that they rely on the assumption that individual patterns are equivalent or at least comparable. However, a collocata analysis has shown that these expressions are not equally representative of the insolence pragmeme (although it must be said that the accuracy of a collocata analysis also depends on the availability of data).

In any case, from a semantic pragmatic standpoint, it is important to note that these variations of the pattern are not all synonymous and represent a metapragmatic comment

on attitude which ranges from positive evaluation (courage) to negative evaluation (impudence). The readings of the evaluations are, however, not readily available outside of any context, although some tend to be more typically associated with negative evaluation; the ambivalence of the reading appears to be part of the construction itself. The collocative analysis provides some qualitative clues as to the specific usage of the individual patterns, and their specific inferences. Ambivalence appears to be a factor in the differentiation of the patterns. *Have the nerve* is notably extremely variable in terms of value assigned to the metapragmatic comment, which ranges from laudatory to accusatory.

On the other hand, the surprisingly low-key *have the cheek* appears deceptively insignificant in the graph and the data set. However, it can be argued that *have the cheek* is a well-entrenched formulaic expression despite its relative infrequency compared to other fillers. The entrenchment value, or prototypical value of the expression in the sense of ‘insolent’, is a measure that remains difficult to assess purely from a usage-based perspective. If the consistency of the semantic and pragmatic reading has an effect on the representativeness of that expression within the conceptual space, then, despite its low frequency, *cheek* (and *cheekiness*) is likely a central representative of a certain subtype of insolence characterized by a positive evaluation as shown in Smith (2021).

The signs of extensibility of *have the N* (body part) are present, although not major with three new body part constructions appearing in the 1900s including *guts*, *stomach* and *balls*. The extension of the pattern to more dysphemistic body part names is likely related to the greater proportion of colloquial oral discourse in recent corpora. We may note that the emergence of *have the stones* and *have the cojones* is likely triggered via synonymy with *balls*. *Stones* has an extremely low frequency rate in the COHA with a total of 3 occurrences ranging from 1980 to 2010; *cojones* appears once in the COHA in a TV/Movie dialogue line from [1991] in (27):

(27) Now all you have to do is have the **cojones** to say it to your boss instead of his secretary. (1991. *Suburban Commando*, TV Movie)

Yet again this low frequency cannot be equated with low productivity but rather is a sign of innovation. The much larger OEC gives a relative frequency of 0.02 for *have the stones* (28) showing that it is far from insignificant; *have the cojones* has a frequency of 0.01.

(28) In all honesty, I did not think that Jack McConnell and his cohorts would have the **cojones** to take on Scotland’s suicidal, macho fag-and-booze culture, so congratulations to them (5 February 2006. *Scotland on Sunday*.)

We believe that these remarks and issues point to the need to pursue in an depth low-level study of micro constructions from a comparative perspective.

4. Discussion of Onomasiological Productivity and “Conceptual Space”

We now discuss our findings concerning HAVE THE N (BODY PART/ATTITUDE) TO ↔ METAPRAGMATIC COMMENT, specifically regarding the issue of how to overcome methodological issues relating to a qualitative approach to the constructional architecture and the relative productivity of constructions. The corpus-driven methodology attempted to study a low-level construction belonging to a higher-level pragmatic construction (insolence/(im)politeness) and represents a metapragmatic discourse comment on a co-speaker’s behavior or attitude.

The analysis of the occurrences of the pattern in the COHA showed a high frequency of certain concept words in structures that have become phraseological patterns (*have the right to*), resembling modal comments.

Our question was whether an onomasiological viewpoint would be able to assess the extensibility capacity of a construction; the study has shown that extensibility is present in terms of the inclusion of new body part nouns to the pattern in the 1970s such as *guts*, *stomach* and *balls* which follow the metaphorical pathway of BODY PART IS COURAGE or BODY PART-IS INSOLENCE. However, beyond this observation, it is difficult to assess what proportion of the onomasiological space these constructions take. The method applied

here showed the intricacies of identifying patterns with respect to semantic, pragmatic and conceptual criteria, and also revealed its limitations. We used raw frequencies, relative frequencies and collocate behaviors to trace *have the N to* patterns over time in the COHA. The wide variety of fillers in the N position were categorized into broad conceptual categories (emotions and attitudes) and body part noun fillers were compared to conceptual noun fillers. The tools used included raw frequencies, relative frequencies and collocate patterns to establish a preliminary view of the diachronic evolution of *have the N to*. The results show that what is required is a more cohesive picture of the onomasiological landscape of the politeness metapragmatic comment superconstruction.

We argue that the results confirm that frequency parameters alone do not adequately show the significance of a pattern on a cognitive level, as is illustrated by the poor data for *have the cheek*, despite the high entrenchment (that is the high degree of mental correlation between cheek–insolence) of this pattern in the lexicon of speakers. We believe this shows the importance of continuing this line of reasoning and testing a larger number of subconstructions within a network of superconstructions. If we follow Goldberg (2019, p. 67), who underlines that studies have shown that “clusters with higher density tended to attract new members, just as the notion of coverage predicts”, then the density of body part fillers associated with insolence/(im)politeness points to a high coverage of conceptual space.

Although this study focused on a very low-level construction, we argue that studying more low-level lexico-grammatical patterns can provide a more reliable understanding of the emergence and development of a family of constructions. If we look at grammaticalization and constructional change, increasing intersubjectivity (expression of approval or disapproval and indignation) can be expressed by low-level lexicogrammatical patterns which in turn point to the relevance of carrying out corpus-driven studies of expressions that are not fully grammaticalized. The expression of modality and intersubjectivity is far from being the purview of grammar only, which therefore continues to call into question the nature of the lexicon-grammar interface. We can only concur with Goldberg’s (2019, p. 146) desire “to capture relational meanings or the myriad ways in which words and grammatical constructions combine to give rise to interpretations in context”. To contribute to achieving this goal, a more in-depth study of collocational behavior in diachronic corpora (including corpora representing periods preceding 1810, such as the English Historical Book Collection or EHBC) is required. Furthermore, the identification of the family of constructions belonging to the insolence pragmeme will also provide a more accurate assessment of shift and productivity from the onomasiological or conceptual standpoint. Ultimately, we aim to achieve a better understanding of the organization of the constructional families and the nature of the constructional families.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data is freely available using the COHA.

Acknowledgments: Many thanks to the expert reviewers and editing team who helped to considerably improve on earlier versions of this paper. All remaining errors are mine entirely.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ “Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.” (Goldberg 2006, p. 5) Note, however, that Goldberg (2019) has since criticized this reference to “sufficient frequency” as “nonsensical”.
- ² “Changes in form and meaning of a construction can be studied through frequency measurements of its variants: An important concept in the present work is thus the idea that constructions are not fixed, but flexible, displaying formal and functional variation.” (Hilpert 2013, p. 6).

References

Archival Sources

The Oxford English Dictionary, Oxford: Oxford University Press. Available online: <https://www.oed.com/> (accessed on 9 January 2023).

The Oxford English Corpus, Oxford: Oxford University Press. Accessed via Sketch Engine. On 9 January 2023.

The Sketch Engine. Corpus management platform available online at <http://www.sketchengine.eu> (accessed on 9 January 2023).

Published Sources

- Allan, Keith. 2016. The pragmeme of insult. In *Pragmeme and Theories of Language Use*. Edited by Keith Allan, Alessandro Capone and Istvan Kecskes. Berlin and Heidelberg: Springer, pp. 69–84.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.
- Audring, Jenny, Booij Geert, and Jackendoff Ray. 2017. Menscheln, kibbelen, sparkle: Verbal diminutives between grammar and lexicon. In *Linguistics in the Netherlands 2017*. Edited by Bert Le Bruyn and Sander Lestrade. Amsterdam: Benjamins, pp. 1–15.
- Booij, Geert, and Jenny Audring. 2018. Partial Motivation, Multiple Motivation: The Role of Output Schemas in Morphology. In *The Construction of Words. Advances in Construction Morphology*. Edited by Geert Booij. Dordrecht: Springer, pp. 59–80.
- Bousfield, Derek. 2008. *Impoliteness in Interaction*. Amsterdam and Philadelphia: John Benjamins.
- Budts, Sara, and Peter Petr . 2020. Putting connections centre stage in diachronic construction grammar. In *Nodes and Networks in Diachronic Construction Grammar*. Edited by Lotte Sommerer and Elena Smirmova. Amsterdam and Philadelphia: John Benjamins, pp. 317–51.
- Bybee, Joan Lea. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Davies, Mark. 2010. The Corpus of Historical American English (COHA): 400 Million Words, 1810–2009. Available online: <https://www.english-corpora.org/coha/> (accessed on 9 January 2023).
- De Smet, Hendrik, Frauke D’hoedt, Lauren Fonteyn, and Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29: 197–234. [CrossRef]
- Durkin, Philip. 2016. The OED and HTOED as tools in practical research: A test case examining the impact of loan words on areas of the core lexicon. In *The Cambridge Handbook of English Historical Linguistics*. Edited by K yto Merja. Cambridge: Cambridge University Press, pp. 390–406.
- Fernandez-Dom nguez, Jes s. 2019. Onomasiological approach. In *Oxford Research Encyclopedia of Linguistics*. Edited by Mark Aronoff. Oxford: Oxford University Press, pp. 1–28.
- Garmendia, Joana. 2018. *Irony*. (Key Topics in Semantics and Pragmatics, p. 1). Cambridge: Cambridge University Press.
- Geeraerts, Dirk. 2016. Entrenchment as Onomasiological Salience. In *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Edited by Hans-J rg Schmid. Berlin: De Gruyter, pp. 153–74.
- Ghesqu re, Lobke, Brems Lieselotte, and Freek Van de Velde. 2015. Intersubjectivity and intersubjectification Typology and operationalization. *English Text Construction* 5: 128–52. [CrossRef]
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. 2016. Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. *Language and Cognition* 8: 369–90. [CrossRef]
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Gries, Steven T. H., and Anatol Stefanowitch. 2004. Extending colostruational analysis: A corpus-based perspectives on ‘alternations’. *International Journal of Corpus Linguistics* 9: 97–129. [CrossRef]
- Gyselincx, Emmeline. 2020. (Re) Shaping the constructional network. In *Nodes and Networks in Diachronic Construction Grammar*. Edited by Lotte Sommerer and Elena Smirmova. Amsterdam and Philadelphia: John Benjamins, pp. 107–40.
- Heine, Bernd, and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2006. Keeping an eye on the data: Metonymies and their patterns. In *Corpora in Cognitive Linguistics: Metaphor and Metonymy*. Edited by Anatol Stefanowitsch and Stephen Gries. Berlin: Mouton de Gruyter, pp. 120–46.
- Hilpert, Martin. 2013. *Constructional Change in English. Developments in Allomorphy, Word Formation and Syntax*. Cambridge: Cambridge University Press.
- Hilpert, Martin, and Samuel Bourgeois. 2020. Intersubjectification in constructional change: From confrontation to solidarity in the “Sarcastic much?” construction. *Constructions and Frames* 12: 96–120. [CrossRef]
- Hoffmann, Thomas. 2020. Construction grammar and creativity: Evolution, psychology, and cognitive science. *Cognitive Semiotics* 13: 20202018. [CrossRef]
- Imo, Wolfgang. 2015. Interactional Construction Grammar. *Linguistics Vanguard* 1: 69–77. [CrossRef]
- Killgarriff, Adam, Pavel Rychl y, Pavel Smr z, and David Tugwell. 2004. The sketch engine. Paper presented at 11th EURALEX International Congress, Lorient, France, July 6–10. Edited by Geoffrey Williams and Sandra Vessier. Lorient: Universit  de Bretagne-Sud, Facult  des Lettres et des Sciences Humaines, pp. 105–16.
- Lehmann, Claudia. 2023. Multimodal markers of irony in televised discourse. *A Corpus-Based Approach. Multimodal Im/Politeness: Signed, Spoken, Written* 333: 251–72.

- Lorenz, David. 2020. Converging variations and the emergence of horizontal links. In *Nodes and Networks in Diachronic Construction Grammar*. Edited by Lotte Sommerer and Elena Smirnova. Amsterdam and Philadelphia: John Benjamins, pp. 243–76.
- Masini, Francesca, and Jenny Audring. 2019. Construction morphology. In *The Oxford Handbook of Morphological Theory*. Edited by Jenny Audring and Francesca Masini. Oxford: Oxford University Press, pp. 365–89.
- Mey, Jacob L. 2001. *Pragmatics: An Introduction*, 2nd ed. Oxford and Malden: Blackwell. First published 1993.
- Mey, Jacob L. 2010. Reference and the pragmeme. *Journal of Pragmatics* 42: 2883–88. [CrossRef]
- Niemeier, Susanne. 2000. *Straight from the Heart—Metonymic and Metaphorical Explorations. Metaphor and. Metonymy at the Crossroads. A Cognitive Perspective*. Edited by Antonio Barcelona. Berlin and New York: Mouton de Gruyter, pp. 195–213.
- Perek, Florent. 2020. Productivity and schematicity in constructional change. In *Nodes and Networks in Diachronic Construction Grammar*. Edited by Lotte Sommerer and Elena Smirnova. Amsterdam and Philadelphia: John Benjamins, pp. 141–66.
- Petré, Peter. 2019. How constructions are born. The role of patterns in the constructionalization of be going to INF. In *Patterns in Language and Linguistics: New Perspectives on a Ubiquitous Concept (Topics in English 104)*. Edited by Beatrix Busse and Ruth Moehlig-Falke. Berlin: Mouton de Gruyter, pp. 157–92.
- Smith, Chris A. 2021. Approche constructionnelle de l'émergence d'expressions d'insolence en anglais. *Cahiers de lexicologie* 119: 235–64.
- Traugott, Elizabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford Studies in Diachronic and Historical Linguistics. Oxford: Oxford University Press.
- Traugott, Elizabeth Closs. 2015. Towards a coherent account of grammatical constructionalization. In *Diachronic Construction Grammar*. Edited by Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer and Spike Gildea. Amsterdam: John Benjamins, pp. 51–80.
- Traugott, Elizabeth Closs. 2019. Constructional pattern-development in language change. In *Patterns in Language and Linguistics: New Perspectives on a Ubiquitous Concept (Topics in English 104)*. Edited by Beatrix Busse and Ruth Moehlig-Falke. Berlin: Mouton de Gruyter, pp. 125–56.
- Wray, Alison. 2017. Formulaic Sequences as a Regulatory Mechanism for Cognitive Perturbations During the Achievement of Social Goals. *Topics in Cognitive Science* 2017: 1–19. [CrossRef] [PubMed]
- Zenner, Eline, Dirk Speelman, and Dirk Geeraerts. 2014. Core Vocabulary, Borrowability and Entrenchment: A Usage-Based Onomasiological Approach. *Diachronica* 31: 74–105. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Clitic Placement and the Grammaticalization of the Future and the Conditional in Old Catalan

Aina Torres-Latorre ^{1,*} and Andreu Sentí ^{1,2}¹ Departament de Filologia Catalana, Universitat de València, 46010 València, Spain² Institut Interuniversitari de Filologia Valenciana, Universitat de València, 46010 Valencia, Spain; andreu.senti@uv.es

* Correspondence: aitola@alumni.uv.es

Abstract: The romance future and conditional tenses are the result of the grammaticalization of Latin periphrasis, mainly CANTĀRE HABEŌ. In some medieval Romance languages, including Catalan, two types of forms existed: synthetic forms (*faré* ‘I will do’) and analytical forms (*fer-lo he* ‘I will do it’). Analytical forms do not present univerbation and are thus less grammaticalized than synthetic forms. The present work aims to study the distribution of synthetic and analytical forms diachronically. A diachronic corpus (11th c.–16th c.) was compiled to serve this purpose. According to the syntactic restrictions of clitic placement, analytical forms could appear in the same syntactic environments than synthetic forms with postverbal pronouns (*faré-lo* ‘I will do it’). Therefore, only those contexts are analysed to assess the degree of grammaticalization. Some recent works point out that the grammaticalization of future and conditional was more advanced in the eastern languages of the Iberian Peninsula, such as Catalan, than in the western ones. The results from our corpus confirm these differences. In addition, the data show another grammaticalization process: the evolution of clitic placement towards a fixed preverbal position.

Keywords: clitic placement; future; conditional; grammaticalization; univerbation; Old Catalan

Citation: Torres-Latorre, Aina, and Andreu Sentí. 2023. Clitic Placement and the Grammaticalization of the Future and the Conditional in Old Catalan. *Languages* 8: 182. <https://doi.org/10.3390/languages8030182>

Academic Editors: Renata Enghels, Marlies Jansengers and Tom Bossuyt

Received: 2 February 2023

Revised: 23 June 2023

Accepted: 7 July 2023

Published: 28 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction¹

The study of clitic placement in Old Romance languages is of great importance as regards the understanding of the formation of two verbal tenses: the future and the conditional (FC). These tenses were the result of the grammaticalization of a Vulgar Latin periphrasis, CANTĀRE HABEŌ (‘I have to sing’), in most Romance languages, including Catalan (cf. Coleman 1971; Fleishman 1982; Lehmann 1995; Pérez Saldanya 1998; Hopper and Traugott 2003; Company 2012). Nevertheless, this grammaticalization was not completed in some Romance varieties: the so-called synthetic future or conditional (SFC) (1a) coexisted beside the analytical future or conditional (AFC), which was possible only with mesoclisism, i.e., with the presence of a clitic pronoun between the infinitive and the auxiliary² (1b) (Castillo Lluch 1996, 2002; Eberenz 1991; Moll [1952] 2006; Badia i Margarit 1984):

- (1) a. si Déus li ajudarà (*Costums*: 490)
if God him.DAT help.FUT.3SG
‘if God will help him’
- b. e Déus ajudar -li à (*Desclot, Crònica*: III.85)
and God help.INF him.DAT have.AUX.3SG
‘and God will help him’

This formal variation in the FC shows two different stages from the grammaticalization of the original Latin periphrasis. The SFC construction attained univerbation, and the infinitive and auxiliary, therefore, lost their boundaries and coalescence took place (cf. Lehmann 1985, 2020).³ The AFC construction is, on the contrary, a less grammaticalized

form, since the infinitive and the auxiliary are separated by at least a clitic pronoun. In the Middle Ages, this analytic construction could be found in western Romance varieties, such as Galician-Portuguese (2a), Astur-Leonese (2b), Spanish (2c), Aragonese (2d), Catalan (1b), Occitan (2e), and some Italian dialects (cf. Pérez Saldanya 1998; Paden 1998, pp. 183–86; de Andrés Díaz 2013, pp. 496–97, 499–500; Company 2006; Buridant 2019, pp. 371–72; Primerano and Bouzouita 2023, among others). Nowadays, it can only be found in Portuguese, in which it is preserved in formal registers (Cunha and Cintra 1992, p. 281).

- (2) a. Et eu dar vos ey este caualo
 and I give.INF you.PL.DAT have.AUX.1SG this horse
 ‘And I will give you this horse’
 (apud Martins 1994, p. 159)
- b. A ti fazer te lo é
 to you.SG.DAT make.INF you.SG.DAT it.ACC have.AUX.1SG
 uerdat
 true
 ‘To you I will make it come true’
 (apud Primerano et al. 2022)
- c. E tornar los é a sus terras e
 and bring back.INF they.ACC have.AUX.1SG to their lands and
 a sus heredades [...] to their landed properties
 ‘And I will bring them back to their lands and their landed properties’
 (apud Bouzouita 2016: 270)
- d. Dizir uos hemos quoaes horas
 tell.INF you.PL.DAT have.AUX.1PL which hours
 deue auer el infançon
 must.3SG have.INF the nobleman (minor nobility)
 ‘I will tell you which hours the nobleman must have’
 (apud Primerano 2019: 117)
- e. mas servir l’ ai dos ans o tres
 but serve.INF him.DAT have.AUX.1SG two years or three
 ‘but I will serve him two years or three’
 (apud Jensen 1994, p. 243)

The variation between the SFC and the AFC is not accidental. However, there is disagreement between the various proposals, which mainly focus on Old Castilian and whose hypotheses are different. On the one hand, for some scholars, the AFC displays a pragmatic function in which the clitic seems to reinforce the presence of an important entity in the discourse, since it appears in a central position in the predicate (Company 2006, p. 383). Therefore, the AFC is seen as a topicalization strategy (Company 1985, 2006; Company and Medina 1999). This pragmatic approach explains the loss of the AFC by means of the pragmatic restrictions of the form, which is considered marked, in contrast with the SFC, a less restricted construction that progressively starts to cover the space of AFC (Company 2006, p. 410). According to this view, the syntactic restriction of clitic placement in Old Romance languages do not play a role in the variation between SFC and AFC.

On the other hand, other scholars have precisely highlighted the connection between clitic placement and the distribution of the different forms of FC. In the Middle Ages, clitic placement with the FC is governed by the same syntactic–pragmatic restrictions than those with other tenses (Eberenz 1991; Castillo Lluch 1996, 2002; Bouzouita 2011, 2012, 2016; Sentí and Bouzouita 2022). The three following points support this perspective against the pragmatic approach and the perception of AFC as a marked configuration: (i) the AFC and SFC with postverbal clitics (SFC-p, also known as postposition or enclisis) can exclusively appear in the syntactic environments that require a postverbal clitic; (ii) in some linguistic varieties, such as Old Castilian, the AFC is (almost) the only available construction for

the environments which demand postposition; (iii) the SFC with preverbal clitics (SFC-a, also known as anteposition or proclisis) are permitted in those syntactic contexts in which preverbal clitics are expected with other verbal tenses. The syntactic–pragmatic restrictions of clitic placement will be discussed in more detail in Section 1.1.

Additionally, some authors have studied the AFC as a modal periphrasis with a marked configuration, a special case of fronting (cf. Fischer 2005; Mathieu 2006; Elvira 2015). Octavio de Toledo (2015), for instance, relates the AFC with fronting configurations of modal periphrasis in Spanish (i.e., *hacer lo debo* ‘I must do it’), whereby the infinitive and the clitics of the AFC display an informative function, which results in a weak focus (cf. Batllori and Hernanz 2015). Thus, although AFC and SFC-p appear in the same syntactic environments, both constructions would not share the same informative function. In her formal analysis, Batllori (2015) argues that the AFC expresses different modal, evidential and speech act meanings due to the movement from S_v to the left periphery, into Mood or Modality nodes. According to these authors, the loss of AFC (and SFC-p) is not because of the increase in SFC-a, but because of a more general linguistic change in Old Spanish: the various information structure options that were coded using word order fade away and they are replaced by syntactic marking (Octavio de Toledo 2015, p. 210). Although these proposals offer an explanation of the AFC and SFC-p variation, we do not consider the AFC as a marked configuration. As has been underlined before, in Old Castilian the AFC is (almost) the only possible construction in some syntactic environments. Modal or informative differences would thus not be expected in this form as no other form of FC takes up its place. Furthermore, in this paper, the AFC is not considered a periphrasis because, among other reasons, modal meanings have not been attested in our data.⁴ In addition, important differences have been attested in the evolution of these forms between some Romance languages (i.e., the Catalan data of this paper, see Section 3, in contrast with Old Castilian). In fact, we aim to provide empirical data in favour of a language contact hypothesis (Bouzouita 2016) to explain the diatopic propagation of the SFC (Primerano et al. 2023; see Section 1.2).

1.1. Clitic Placement in Old Catalan

The position of pronominal clitics with respect to the verb in Old Catalan (and Old Romance languages) is motivated by syntactic–pragmatic restrictions (Fischer 2002; Batllori et al. 2005; Sentí and Bouzouita 2022)—unlike in the present-day language, in which morphosyntactic restrictions related to the type of verb form govern clitic positioning (cf. GIEC 2016, Section 8.3.3; Bonet 2002).⁵

The syntactic–pragmatic environments in which the finite verb appears determine the position of clitics. Clitics generally occur in proclisis in subordinate clauses, while there are three possibilities in main clauses: (i) clitics can be proclitic in certain syntactic contexts, such as negative sentences, (ii) they appear only enclitically in others, and (iii) they admit variation in cases such as when the verb follows the subject. According to the literature, and particularly studies on Old Castilian, the following three groups of syntactic environments that determine the position of pronominal clitics have been established (cf. Eberenz 1991; Castillo Lluç 1996, 2002; Bouzouita 2008a, 2008b, 2011, 2012, 2013, 2016). The three groups are detailed in Table 1 and will be explained below.

Group 1 always leads to proclisis. Certain preverbal constituents always trigger proclisis, such as a *wh*-word, a negation marker, a preverbal NP non-coreferential with the clitic, a preverbal predicative complement, the contrastive conjunction *ans* (‘but’) or an infinitive or a participle dislocated before the verb (fronting).

Group 2 has only postverbal clitics. The syntactic–pragmatic environments in which a non-preverbal clitic is expected in all verb tenses are those that appear with the verb in the first position, paratactic clauses, after a vocative, and in coordinated adversative clauses after the conjunction *mas* or *però* (‘but’).

Finally, Group 3 is made up of those environments in which variation in clitic placement occurs: following a preverbal subject, after a dislocated preverbal NP coreferential

with the clitic pronoun, after a preverbal adverbial phrase, following a coordinative or disjunctive conjunction and following a non-root or absolute clause. The default position in these syntactic contexts is enclisis. However, some (pragmatic) marked configurations can trigger proclisis. Preverbal clitics appear when the preverbal constituent proves to be emphatic or prominent in the information structure. Following Granberg (1999), the emphatic constituent is considered focus if it receives informative prominence in the discourse.⁶ Even though analysing information structure in written medieval texts is not always reliable, some discourse phenomena can indicate the presence of emphasis (for example, a contrastive subject regarding a different previous entity in discourse). Regarding clitic placement, some studies on present-day Galician (Álvarez Blanco et al. 1986) and Old Castilian (Granberg 1988, 1999; Castillo Lluch 1996; Bouzouita 2008a, 2008b, 2011, 2012, 2016) have observed that this pragmatic function (emphasis) accounts for the presence of preverbal clitics in the aforementioned syntactic environments. Similar behaviour has also been recently observed in Old Catalan (Bouzouita and Sentí 2022; Sentí and Bouzouita 2022; Torres-Latorre 2023). Even so, a detailed exploration of the consequences of information structure in clitic placement in Old Romance languages is still pending.

Table 1. Syntactic–pragmatic environments governing clitic placement.

| Syntactic–Pragmatic Environment | | Preverbal Clitics | Postverbal Clitics |
|---------------------------------|---|-------------------|--------------------|
| Main clauses: Group 1 | <i>Wh</i> -pronoun | * | |
| | Negation | * | |
| | Non-coreferential NP | * | |
| | Prepositional complement | * | |
| | Predicative complement | * | |
| | Contrastive <i>ans</i> ‘but’ | * | |
| | Infinitive | * | |
| Participle | * | | |
| Main clauses: Group 2 | P1 (verb in first position) | | * |
| | Paratactic clause | | * |
| | Vocative | | * |
| | Contrastive <i>mas, però</i> ‘but’ | | * |
| Main clauses: Group 3 | Subject | * | * |
| | Coreferential NP | * | * |
| | Adverbial | * | * |
| | Coordinative conj. <i>e, y, i</i> ‘and’ | * | * |
| | Disjunctive conj. <i>o</i> ‘or’ | * | * |
| | Non-root and absolute clause | * | * |

With regard to the grammaticalization of the FC, and according to the above-mentioned syntactic restrictions of clitic placement, the AFC could appear in the same syntactic contexts as the SFC with postverbal pronouns (Group 2 and 3). These contexts will, therefore, be examined in this work.

1.2. Objectives

According to the summarised state of the art, two different processes of linguistic change related to clitic placement are dealt with in this paper. First, the variation permitted between a more grammaticalized construction, i.e., SFC-p, and a less grammaticalized version of the FC, i.e., the AFC, since these forms appear in the same syntactic contexts. As previous studies have shown, both constructions are frequent in Old Catalan (Bouzouita and Sentí 2022; Sentí and Bouzouita 2022; Torres-Latorre 2023). However, this was not the case in Old Castilian. As already urged by Bouzouita (2016), SFC-p should be scrutinised in more detail in Ibero-Romance, because, whereas this form was rarely found in Old Castilian texts, it was quite a lot more frequent in some eastern texts written in this language, as already noted by Saralegui (1985). Some recent works (Bouzouita 2016; Bouzouita and Sentí 2022; Primerano et al. 2022) point out that the synthetic forms could have been spread

from Catalan to other languages and varieties in the Iberian Peninsula during the 11th–13th centuries (from Catalan to Navarro-Aragonese, from Navarro-Aragonese to Castilian, etc.), signifying that language contact may have played an important role. The empirical findings of the grammaticalization of the FC in Old Catalan, therefore, prove crucial to understand the Old Ibero-Romance.

The objective of this paper is to study the distribution of the SFC and the AFC diachronically and quantitatively and to analyse the grammaticalization process from the 11th to the 16th century. In this respect, an increase in the frequency of SFC-p, with univerbation and a more grammaticalized form, and a consequent decrease in the AFC would be expected, bearing in mind that the AFC is no longer used in contemporary Catalan. In fact, this development has been noted in Old Castilian and in Old Navarro-Aragonese in the 13th and 14th centuries. In Old Castilian, SFC-p was almost non-existent in the 13th century, and it remained uncommon in the 14th century, but the increase in its presence appears to be statistically significant (Primerano et al. 2023, Section 4.1). Although SFC-p was not as rare in Old Navarro-Aragonese as it was in Old Spanish, an increase in its use is also attested in these two centuries (Primerano and Bouzouita 2023).

Furthermore, another grammaticalization process participated in the evolution of Old Catalan: the evolution of clitic placement from the above-mentioned syntactic distribution towards a fixed preverbal position (as in present-day Catalan). In spite of the previous knowledge on pronominal clitics in Old Catalan (Fischer 2002; Batllori et al. 2005; Bouzouita and Sentí 2022; Sentí and Bouzouita 2022), this topic has yet to be studied quantitatively and qualitatively. There is lack of knowledge on which variation contexts favour preverbal placement and on the extent to which proclisis is linked to emphasis (or whether it can also appear in a neutral, i.e., non-emphatic environment, and therefore be used in innovative contexts). The same applies to its evolution throughout history and the fixing of word order with the clitics in proclisis to the finite verb, or its dialectal and discursive variations. This is the second purpose of this paper: we wish to contribute to explaining this process with data regarding the FC, again from the 11th to the 16th century.

In the following section (Section 2) we explain our selected corpus and some methodological issues. The empirical results and a discussion of the corpus-based study are then addressed: Section 3 shows the distribution of the AFC and SFC-p in Old Catalan, while Section 4 summarises the evolution of clitic placement towards anteposition. Finally, we present our conclusions in Section 5.

2. Methodology and Corpus

Old Catalan usually covers the period between the first attested Catalan texts (11th century) and the beginning of the modern age (16th c.) (Martines and Pérez Saldanya 2009).⁷ In order to carry out the study of the grammaticalization of the FC and clitic placement in this period, we have used the *Corpus Informatitzat del Català Antic* (CICA 2006) (Claveria and Torruela 2012), which is the only available corpus that is devoted to the study of this period. Although the CICA (2006) corpus is not a morpho-syntactically annotated corpus, the extraction of the FC with clitics has been carried out semi-automatically by seeking the forms through a search for verb endings. Moreover, the analytic forms were found thanks to the diacritic accentuation available in the electronic edition of the texts.

The high token frequency of the FC made it necessary to restrict ourselves to select some texts from the CICA (2006) for this study.⁸ We have selected only those texts that were written in the same period as the preserved manuscript. We have selected texts from the second half of each century in order to analyse the linguistic change from one period to another. Diatopic variation has also been taken into account: we have established a balanced subcorpus for each group of Catalan dialects with a similar number of tokens (eastern and western Catalan). Despite having paid attention to the extraction of a similar number of tokens for each dialect, it was not easy to obtain these results for all the periods considered (see Table 2).⁹ Finally, a reasonable range of textual typologies has also been considered, such as religious, legal, narrative, historiographic texts and letters, although

attempting to preserve the representativeness of each discursive tradition in each period was quite challenging (Kabatek 2005, 2013).

Table 2. FC tokens retrieved for each period in each dialectal group.

| Dialect | 11th–12th | 13th | 14th | 15th | 16th |
|--------------------------|-----------|------|------|------|------|
| Eastern Catalan | - | 732 | 925 | 213 | 360 |
| Western Catalan | - | 991 | 333 | 953 | 357 |
| No dialectal adscription | 218 | 1186 | 28 | - | - |
| Total | 218 | 2909 | 1286 | 1166 | 717 |

Those examples of the SFC and AFC with adverbial clitic pronouns (e.g., *en* < INDE and *hi* < IBI) have been analysed separately. This decision made it easier to compare Catalan and other Ibero-Romance languages that barely conserve these pronouns (e.g., Primerano 2021 for Navarro-Aragonese). Moreover, it also allowed a comparison between the behaviour of personal clitic pronouns and that of adverbial clitic pronouns, which has proved relevant (Section 3.1).

3. The Distribution of AFC and SFC-p in Old Catalan

The analytical and synthetic forms in Old Romance do not always coexist for pragmatic–syntactic reasons (Section 1). In fact, the AFC and SFC with postverbal pronouns appear in the same pragmatic–syntactic environments, whereas the SFC with preverbal pronouns are used in other contexts. Thus, the degree of grammaticalization of the FC can be observed only by means of the variation between the AFC and SFC-p. The use of SFC-p rather than the AFC shows a more grammaticalized construction of these two verbal tenses, since synthetic forms display univerbation (Lehmann 1985, 2020). Previous studies have already attested SFC-p from the 13th c. in Old Catalan (Bouzouita and Sentí 2022; Sentí and Bouzouita 2022; Torres-Latorre 2023). Although the AFC is more frequent, the presence of SFC-p in early periods of the language shows a high degree of grammaticalization. In fact, the aforementioned studies have found SFC-p to be more frequent in Old Catalan than in Old Navarro-Aragonese and in Old Spanish.

The empirical results regarding the distribution of the AFC and SFC-p (with personal pronouns) in Old Catalan are summarised in Table 3. This table considers only two of the three possible constructions of the FC: the AFC and SFC-p. The data contained in it, therefore, originate from the syntactic environments of Groups 2 and 3 explained previously (Section 1.1). The results with adverbial pronouns (*en*, *hi*) will be presented hereafter (Section 3.1).

Table 3. The distribution of the AFC and SFC-p with personal pronouns (11th–16th c.).

| Chronology | AFC | SFC-p |
|------------------------------------|-----------------|-----------------|
| Preliterary Catalan (11th–13th c.) | 50% (1/2) | 50% (1/2) |
| 13th century, 2nd half | 61.2% (180/294) | 38.8% (114/294) |
| 14th century, 2nd half | 60.2% (50/83) | 39.8% (33/83) |
| 15th century, 2nd half | 58.1% (25/43) | 41.9% (18/43) |
| 16th century, 2nd half | 80% (12/15) | 20% (3/15) |
| Total | 61.3% (268/437) | 38.7% (169/437) |

At first sight, both forms, the AFC and SFC-p, are present throughout the period. The distribution of the two forms does not have the exact expected evolution, signifying that SFC-p does not appear to gain ground, while the AFC disappears over time. The distribution of the AFC and SFC-p does not change in the following periods. Leaving aside preliterary Catalan, which is deemed unusable since only two tokens have been attested,¹⁰ the differences between the remaining centuries shown in Table 3 are not statistically significant ($\chi^2 = 2.4328$, $p = 0.487565$; significant at $p < 0.05$). The AFC always predominates, with

61.3% (268/437) of the total of instances, over SFC-p, with 38.7% (169/437). There appears to be a preference for the AFC in the 16th century (80%, 12/15), but the limited number of both constructions in this century makes the statistical differences between periods non-significant. It would, therefore, appear that the more grammaticalized construction (SFC-p) does not increase in frequency over time, as would be expected in this kind of linguistic change process (Bybee 2003).¹¹

However, some empirical findings may be relevant in order to understand the grammaticalization of the FC. First, it is important to point out that an example of FC with postverbal clitics (SFC-p) is found in the 11th c. (3). The presence of a clitic cluster could have triggered this appearance of the unverbated form, which shows the grammaticalized FC construction at an early date, quite a long time before its appearance in other Ibero-Romance languages. In fact, SFC-p is already a productive construction in the 13th century, considering the 38.8% (114/294) of instances, whereas it is a rare form in Old Castilian, as will be seen later.

(3) et **ajudaré** **·ls** **te** a tener
 and help.FUT.1SG them.ACC you.DAT to have.INF
 contra tots homines et contra totas fēminas (*Documents feudals* 3: 77)
 against all men and against all women
 ‘and I will help you have them against all men and against all women’

Leaving aside preliterary Catalan (see Note 8), the data show a progressive decrease in both the AFC and SFC-p from the 13th c. onwards. Table 4 (below) shows the relative token frequency of the AFC and SFC-p per million words (in relation to the number of words in each period). The AFC represents 215.55 tokens per million words in the 13th century, whereas it represents 36.69 in the 16th century, that is, a seventh of what it was in the 13th century. The decrease in both the AFC and SFC-p, which can already be observed in the 14th, 15th and 16th centuries, is owing to an increase in SFC-a and will be studied in the following section (Section 4). The limited number of tokens in the 16th century is not owing to differences in the size of the corpus. The reason why only 15 tokens are found in our corpus can be explained by the relative frequency, which indicates the general tendency to decrease of both the AFC and SFC-p constructions. There appears to be a preference for the AFC in the century in question (80%, 12/15), but the limited quantity of both constructions in this century make the statistical differences between periods non-significant.

Table 4. The distribution of the AFC and SFC-p with personal pronouns (11th–16th c.) as regards relative frequency (%) per million words.

| Chronology | AFC | SFC-p |
|------------------------------------|---------------------|---------------------|
| Preliterary Catalan (11th–13th c.) | 39.72 (1/25176) | 39.72 (1/25176) |
| 13th century, 2nd half | 212.55 (180/846841) | 134.62 (114/846841) |
| 14th century, 2nd half | 146.84 (50/340517) | 96.91 (33/340517) |
| 15th century, 2nd half | 58.29 (25/428921) | 41.97 (18/428921) |
| 16th century, 2nd half | 36.69 (12/327032) | 9.17 (3/327032) |

The comparison of the empirical data from Old Catalan with Old Castilian and Old Navarro-Aragonese also provides some important findings to understand the grammaticalization of the FC in the Iberian Peninsula. Old Catalan differs from the behaviour of Old Castilian and Old Navarro-Aragonese: despite the fact that the AFC is the main form throughout the period, it is not as frequent as the two other varieties, as summarised in Table 5.

Table 5. AFC and SFC-p in Old Catalan, Old Navarro-Aragonese and Old Castilian (13th–14th c.).

| Language | Chronology | AFC | SFC-p |
|---------------------------------|--------------|-------------------|-----------------|
| Catalan | 13th century | 61.2% (180/294) | 38.8% (114/294) |
| | 14th century | 60.2% (50/83) | 39.8% (33/83) |
| | Total | 61% (230/377) | 39% (147/377) |
| Navarro-Aragonese ¹² | 13th century | 89.1% (49/55) | 10.9% (6/55) |
| | 14th century | 75% (51/68) | 25% (17/68) |
| | Total | 81.3% (100/123) | 18.7% (23/123) |
| Castilian ¹³ | 13th century | 99.8% (1827/1830) | 0.2% (3/1830) |
| | 14th century | 97.3% (359/369) | 2.7% (10/369) |
| | Total | 99.4% (2186/2199) | 0.6% (13/2199) |

In the case of Old Castilian, Primerano et al. (2023, Section 4.1) have discovered that AFC are by far the most frequent in the 13th and the 14th centuries (99.4%) when compared to SFC-p (0.6%); in that of Old Navarro-Aragonese, Primerano and Bouzouita (2023) have found 81.3% of AFC, which is considerably more than the 18.7% obtained for SFC-p in the same two centuries. With regard to Old Catalan, SFC-p represents 38.8% (114/294) and 39.8% (33/83) in the 13th and the 14th centuries, respectively, and it is, therefore, much more frequent. As expected, Old Catalan has the highest degree of grammaticalization of the FC. Furthermore, an increase in SFC-p is perceived in the 14th century in both Old Castilian and Old Navarro-Aragonese, but this is not the case of Old Catalan, in which the variation between the two forms is stable. The 15th century has yet to be studied in these two other languages.

The differences between Catalan on the one hand and Castilian and Navarro-Aragonese on the other can be partially explained by the intervention of certain morphological factors that contribute to the choice of AFC or SFC-p. It has been suggested that certain morphological and even syntactic factors could induce the choice of SFC-p rather than AFC. These factors have been studied principally for Old Castilian (Bouzouita 2016), but also for Old Navarro-Aragonese (Primerano and Bouzouita 2023). Since SFC-p is an uncommon construction in both languages, these works sought the causes of the presence of SFC-p and not of the AFC, which is the construction by default. The inductors of SFC-p would also be triggers of the grammaticalization process of the FC. The factors analysed in these works are the verbal tense (if it is a future or conditional), the conjugation class, the presence of a syncope in the verb and the introduction of a non-finite verbal form after the FC.

A more detailed study of these factors is still necessary for Old Catalan (Torres-Latorre n.d.a). However, some differences between Old Catalan and Old Castilian and Old Navarro-Aragonese can be noted only by observing our preliminary empirical data. First, the presence of a syncope in the FC is much more common in Catalan than in the other two languages. This syncope is the omission of the thematic vowel of the verb and can be found in some verbs of the second conjugation (*volerà* > *volrà* ‘he will want’) and a few verbs of the third conjugation too (*partirà* > *partrà* ‘he will leave’).¹⁴ A verb with a syncope always prefers SFC-p over the AFC (Bouzouita 2016, p. 284; Primerano and Bouzouita 2023), as (4a) shows. That is for phonetic reasons: mesoclisism cannot be conducted if there is a syncope. A form such as **volr-la à* (‘he will want her’) is not possible when following the phonetic rules of these languages, but the same form without a syncope (*voler-la à* ‘he will want her’) is indeed possible. In our data, every instance of non-preverbal clitics with the FC has the SFC-p form, as expected.

During the Middle Ages, the presence of a syncope was optional in the three Romance languages mentioned. Nevertheless, it was much more frequent in Old Catalan. In fact, in contemporary Catalan the presence of a syncope is mandatory in the verbs of the second conjugation which formerly varied (*voler* ‘to want’ > *voldrà* ‘he will want’, with an epenthetic *d*; *saber* ‘to know’ > *sabrà* ‘he will know’, etc.). Contemporary Spanish has, on the contrary, preferred to maintain the omitted vowel in some of the verbs which had an optional

syncope (*saber* ‘to know’ > *sabrà* ‘he will know’, with a syncope; but *perder* ‘to lose’ > *perderà* ‘he will lose’, without a syncope). The differences in frequency in the presence of a syncope in this group of verbs could explain a slight tendency in favour of SFC-p in Old Catalan with respect to the rarity of this construction in Old Castilian and Old Navarro-Aragonese.

Second, there is one particularity of Catalan’s verbal conjugation that distinguishes Old Catalan from Old Castilian and Old Navarro-Aragonese. In Catalan, in the second conjugation, most verbs take a rhizotonic form (Pérez Saldanya 1998, pp. 40–48), that is, they are verbs whose stressed syllable is in the stem, such as *conèixer* (‘to know’, ‘to meet’) or *perdre* (‘to lose’). Alsina (2022, p. 192) finds that rhizotonic verbs also prefer SFC-p over the AFC, but, although a preference for SFC-p is confirmed in our corpus (4b), some tokens of the AFC are also attested (4c). The inclination towards SFC-p is for phonetic reasons: a rhizotonic verb with mesoclisys is not unpronounceable in Old Catalan, but in a form such as *defendre-nos hien* (4c) the main stress remains far from the word ending, in the syllable *-fen-*, and in Catalan, oxytones and words stressed on the penultimate syllable are far more common (Oliva and Serra 2002). This factor could once more entail a slight preference in Catalan for SFC-p when compared with Old Castilian and Old Navarro-Aragonese. However, it is still necessary to quantitatively explore both factors (Torres-Latorre n.d.a).

Moreover, instances of SFC-p in which none of the factors mentioned above (verbal tense, conjugation, syncope, rhizotonic verbs or non-finite verbal forms after the FC) are attested. The first conjugation (verbs with infinitives finished with *-ar*, i.e., *cantar* ‘to sing’) is characterised by a total preference for the AFC in Old Castilian (Bouzouita 2016) and in Old Navarro-Aragonese (Primerano and Bouzouita 2023), but in Old Catalan there are examples such as (4d), in which a verb of the first conjugation appears with the SFC-p construction. The variation between AFC and SFC-p, therefore, requires an explanation. Nevertheless, with regard to the evolution of the variation in the two constructions, these factors appear constant throughout the period studied. For instance, the case of (4d) is from the 13th century, and it does not contain any of the factors that favoured SFC-p. It is possible that there were no cases of SFC-p without the presence of any of these factors in a previous stage of the language, and SFC-p later spread to other verbs, such as *alegrar* from example (4d). Unfortunately, this can only be hypothesised, since previous stages of the Catalan language have not been preserved.

- (4) a. Nostre Senyor no vol la mort del
 Our Lord NEG want.PRES.3SG the death of-the
 peccador, e **volrà-** **la** lo rey? (*Tirant*: 268)
 sinner and want.FUT.3SG it.ACC the king
- ‘Our Lord does not want the death of the sinner, and the King will want it?’
- b. e **conexerets** **-ho** (*Lletres reials*: 564)
 and know.FUT.2PL it.ACC
 ‘And you will know it’
- c. si s crexien de gent, **defender** **-nos**
 if REFL increase.IMP.3PL of people deprive.INF us.ACC
hien la Ciutat (*Fets*: 48r)
 have.AUX.3PL the City
 ‘If they increased in number, they would deprive us of the City’
- d. **Alegrarà** **-s** lo just (*Vides*: 104)
 Be happy.FUT.3SG REFL the just
 ‘The just will be happy’

To sum up, the distribution of the AFC and SFC-p with personal pronouns is stable from the 11th c. to the 16th c., that is, the presence of the AFC does not decrease in favour of SFC-p. Nevertheless, another change occurs in the same period: the grammaticalization process of pronominal clitics and, in particular, their fixation in the preverbal position (Section 4). Both the AFC and SFC-p eventually cease to exist when anteposition becomes the only possible position of clitics. Although signs of evolution towards SFC-p have not been observed in the period studied, the data from Old Catalan show a more advanced

stage of the degree of grammaticalization of the FC in the 13th and the 14th c. than occurs in other peninsular languages. Old Catalan frequently exhibits SFC-p and the AFC is used to a lesser extent when compared with Old Navarro-Aragonese and Old Castilian, as has been shown quantitatively. The morphological differences between the languages (syncope, rhizotonic verbs) could partially justify this contrast, but a more detailed study of these factors is still required (Torres-Latorre n.d.a).

3.1. AFC and SFC-p with Adverbial Pronouns

Thus far, the distribution of the FC with non-adverbial pronouns has been studied. Adverbial pronouns in Old Catalan are mainly *en* (<INDE) and *hi* (<IBI), but also *hic* (<HĪC) (Ribera 2020). The decision to analyse them separately was made for two reasons: (i) *en*, *hi* and *hic* may have different degrees of pronominalization in Old Catalan and could, therefore, behave differently to personal clitic pronouns; (ii) the grammar of Old Romance languages varies in the case of these pronouns, signifying that it is not possible to compare the languages in this respect with the same reliability as in the case with the remaining pronominal system.

The pronouns *en* and *hi* are preserved in contemporary Catalan, unlike in Spanish, from which they disappeared in the 15th c. (Badia i Margarit 1947, p. 128). In the Middle Ages, their behaviour was more clitic-like in Old Catalan than in the other peninsular Romance varieties. In Old Castilian, they are often considered to be pronominal adverbs rather than adverbial pronouns (Badia i Margarit 1947; Meilán García 1994; Polo Cano 2006; Matute 2016). In Old Navarro-Aragonese, they appear to be more similar to clitics than to adverbs (Matute 2016, p. 221). *Hic* appears to be the least pronominalized form of the three and is not conserved in contemporary Catalan. Unfortunately, no instance of the FC with a non-preverbal pronoun *hic* has been attested. Table 6 provides a summary of the data obtained for the AFC and SFC-p with regard to the pronouns *en* and *hi*.

Table 6. The distribution of the AFC and SFC-p with adverbial pronouns (11th–16th c.).

| Chronology | AFC | SFC-p |
|------------------------------------|----------------|----------------|
| Preliterary Catalan (11th–13th c.) | 0% (0/2) | 100% (2/2) |
| 13th century, 2nd half | 45.8% (38/83) | 54.2% (45/83) |
| 14th century, 2nd half | 25% (5/20) | 75% (15/20) |
| 15th century, 2nd half | 28.6% (2/7) | 71.4% (5/7) |
| 16th century, 2nd half | 50% (2/4) | 50% (2/4) |
| Total | 40.5% (47/116) | 59.5% (69/116) |

The lower number of tokens contained in Table 6 in contrast with the total number of occurrences with personal pronouns was to be expected, as only two pronominal clitics were studied. The total number of 437 instances of clitics in the AFC or SFC-p with non-adverbial pronouns (see Table 3) now contrasts with the total of 116 that have been attested. The general distribution of the AFC and SFC-p with adverbial pronouns is almost the opposite of the distribution of the two constructions with non-adverbial pronouns. As a reminder, the AFC represents 61.5% (268/436) of tokens, while SFC-p has 38.5% (168/436) as regards personal clitics. With adverbial pronouns, SFC-p is the main form, with 59.5% (69/116) of tokens, and the AFC follows with 40.5% (47/117). The differences between the two groups of pronouns are statistically significant ($\chi^2 = 16.4142, p = 0.000051$; significant at $p < 0.05$).

With regard to the evolution of the distribution of the AFC and SFC-p with adverbial pronouns, there appears to be some variation, although it is not quantitatively important. In preliterary Catalan, the only two instances attested are SFC-p. However, a greater presence of the AFC has been found in the 16th c., with 50% (2/4), although the total number of instances is highly reduced and it is not, therefore, possible to draw strongly founded conclusions.¹⁵ The differences between the other periods studied are not statistically significant ($\chi^2 = 3.4742, p = 0.324119$; significant at $p < 0.05$), and the AFC does not, therefore,

disappear with adverbial pronouns either. Although there is a preference for SFC-p from the very beginning, its presence does not increase in the period studied. A decrease in the frequency of both constructions is also found with adverbial pronouns, mainly in the 15th and the 16th centuries.

The different status of the pronouns *en* and *hi* may possibly provide an explanation of the difference between the distribution of the AFC and SFC-p. In Old Catalan, adverbial pronouns are less grammaticalized than personal pronouns such as *el* ('him'), *la* ('her') or *li* ('to him, to her'), because the transition from adverbs to pronouns is not yet complete. In fact, the difference between the distribution of the AFC and SFC-p and that of adverbial pronouns has also been observed in Old Navarro-Aragonese (Primerano 2021, p. 418), a language in which adverbial pronouns behave as normal clitic pronouns to an even lesser extent. However, this explanation cannot be completely confirmed without considering the aforementioned morphological and syntactic factors which favour SFC-p rather than AFC (verbal tense, conjugation class, syncope, rhizotonic verbs, non-finite form after the FC). For example, the adverbial pronoun *hi* frequently collocates with the verb *haver* as in *hi haurà* ('there will be') and *haver* has a rhizotonic form with FC (*haure* 'to have'). A review of the factors in relation to this hypothesis is, therefore, also required (Torres-Latorre n.d.a).

In conclusion, the AFC and SFC-p have a stable distribution from the 11th c. to the 16th c. as regards both non-adverbial and adverbial pronouns. There are two differences between Old Catalan, on one hand, and Old Castilian and Old Navarro-Aragonese on the other: (i) the grammaticalization of the FC is more advanced in Old Catalan since SFC-p is found in the above-mentioned syntactic environments in main sentences (Group 2 and Group 3) from the 11th c. onwards, and with an important frequency during the 13th and the 14th c.; (ii) the diverging evolutions of the variation in the distribution of the AFC and SFC-p, which is stable in Catalan rather than in favour of SFC-p. The progressive symmetrical disappearance of AFC and SFC-p is owing to another grammaticalization process, that of pronominal clitics, which is discussed in the following section (Section 4).

4. The Evolution of Clitic Placement towards Anteposition

In the Middle Ages, the position of pronominal clitics with finite verbs was owing to syntactic–pragmatic restrictions (for Old Catalan, see Fischer 2002; Batllori et al. 2005; Senti and Bouzouita 2022). It followed the distribution into three groups of syntactic–pragmatic environments summarised in Table 1 (Section 1.1). However, in present-day Catalan, clitic placement follows only morphological restrictions: (i) finite verbs (apart from the imperative) can be used only with preverbal clitics, (ii) the imperative, the infinitive and the gerund can be used only with postverbal clitics, and (iii) verbal periphrases can be used with either preverbal or postverbal clitics (Bonet 2002, p. 937; GIEC 2016, Section 8, pp. 200–1).

In the case of finite verbs (with the exception of the imperative), clitic placement has, therefore, gone through a fixation process to the preverbal position. This fixation process has not yet been studied sufficiently. In this section, the results obtained for the FC will be analysed by considering the three possible constructions of these two verbal tenses: SFC-a, the AFC and SFC-p. An evolution towards anteposition is expected from the 11th c. to the 16th c. Because of space limitations, only the results obtained for personal pronouns will now be considered and also because adverbial pronouns do not behave differently in the frequencies of SFC-a (Torres-Latorre 2022). The results will be examined first quantitatively and then qualitatively.

Table 7 shows the total number of frequencies of the three constructions, without making any distinction according to the groups of syntactic–pragmatic environments. All the attested instances of the FC with personal pronouns are included. Since the variation between the AFC and SFC-p has already been reviewed, both options are considered together from this point on.

Table 7. The position of personal pronouns with the FC (11th–16th c.).

| Chronology | SFC-a | AFC or SFC-p |
|------------------------------------|-------------------|------------------|
| Preliminary Catalan (11th–13th c.) | 98.7% (151/153) | 1.3% (2/153) |
| 13th century, 2nd half | 86.3% (1854/2148) | 13.7% (294/2148) |
| 14th century, 2nd half | 91.6% (904/987) | 8.4% (83/987) |
| 15th century, 2nd half | 95.4% (889/932) | 4.6% (43/932) |
| 16th century, 2nd half | 97.6% (616/631) | 2.4% (15/631) |
| Total | 91% (4414/4851) | 9% (437/4851) |

At first glance, it will be noted that SFC-a is by far the most frequent throughout the periods studied. Nevertheless, a more detailed analysis shows differences between centuries (statistically significant: $\chi^2 = 126.8946$, $p < 0.00001$; significant at $p < 0.05$). Surprisingly, the period in which the frequency of SFC-a is greatest is that of preliminary Catalan, with 98.7% (151/153) of the instances in our corpus, but methodological issues oblige us not to consider these results (see Note 10). The remaining centuries do show the expected development towards anteposition: the smallest percentage of SFC-a is in the 13th c. (86.3%, 1854/2148), followed by the 14th c. (91.6%, 904/987), the 15th c. (95.4%, 889/932) and finally, the 16th c., in which this percentage reaches 97.6% (616/631) of tokens. The differences among the four centuries are still statistically significant ($\chi^2 = 110.4827$, $p < 0.00001$; significant at $p < 0.05$).

The data from all syntactic-pragmatic environments could be conditioned by a larger or smaller presence of a particular environment in each century, i.e., subordinate clauses, in which the required pronouns are preverbal. The results in Table 7 must, therefore, be viewed with caution. From this point on, attention will consequently be focused on: (i) the main sentence environments in Group 2, i.e., the group in which no token should take the SFC-a form, according to the grammar of the 13th–14th centuries (§Section 4.1), and (ii) the environments in Group 3, which are characterised by the possibility of variation between anteposition and postposition governed by pragmatics (information structure) (Section 4.2).

4.1. The Evolution of Clitic Placement in Group 2

Group 2 is distinguished by the fact that it permits only postverbal clitics (Section 1.1). The expected position of clitics in this environment with the FC would be either the AFC or SFC-p. However, the results from our corpus indicate a different distribution, since SFC-a is possible in some environments from the 13th c. Table 8 provides a quantitative summary of the evolution of the environments in Group 2. The table does not contain preliminary Catalan because no token from this group has been found for this period.

Table 8 proves that anteposition is gaining ground both quantitatively and qualitatively from the 13th century to the 16th century. In Group 2, 0% of this construction was to be expected according to the bibliography for Old Castilian and Old Romance in general (Eberenz 1991; Castillo Llach 1996, 2002; Bouzouita 2008a, 2008b, 2011, 2012, 2013, 2016). In spite of this, examples of preverbal clitics in this group have been attested even in the 13th century. SFC-a is found only after a vocative in the 13th and the 14th centuries (5a); in the 15th century, it appears for the first time with the verb in the first position of the phrase (5b); in the 16th century, SFC-a is also attested in a paratactic clause (5c), with the presence of a contrastive conjunction *mas* or *però* ('but') being the only environment in this group in which SFC-a is never found. Only the context after a vocative, therefore, appears to be sufficiently innovative for the presence of preverbal clitics in the first two centuries, but a clear change can be noted in the 15th, and especially in the 16th, centuries, since other environments in Group 2 have the SFC-a construction.

Table 8. The position of personal pronouns with the FC in Group 2 (13th–16th c.).

| Chronology | Environment | SFC-a | AFC or SFC-p |
|------------------------|------------------------------|--------------|--------------|
| 13th century, 2nd half | P1 | 0% (0/16) | 100% (16/16) |
| | Paratactic | 0% (0/0) | 0% (0/0) |
| | Vocative | 69.2% (9/13) | 30.8% (4/13) |
| | Contrastive <i>mas, però</i> | 0% (0/7) | 100% (7/7) |
| 14th century, 2nd half | P1 | 0% (0/5) | 100% (5/5) |
| | Paratactic | 0% (0/2) | 100% (2/2) |
| | Vocative | 50% (2/4) | 50% (2/4) |
| | Contrastive <i>mas, però</i> | 0% (0/3) | 100% (3/3) |
| 15th century, 2nd half | P1 | 33.3% (2/6) | 66.7% (4/6) |
| | Paratactic | 0% (0/0) | 0% (0/0) |
| | Vocative | 85.7% (6/7) | 14.3% (1/7) |
| | Contrastive <i>mas, però</i> | 0% (0/0) | 0% (0/0) |
| 16th century, 2nd half | P1 | 85.7% (6/7) | 14.3% (1/7) |
| | Paratactic | 100% (1/1) | 0% (0/1) |
| | Vocative | 100% (1/1) | 0% (0/1) |
| | Contrastive <i>mas, però</i> | 0% (0/3) | 100% (3/3) |
| Total | | 36% (27/75) | 64% (48/75) |

- (5) a. E tu, *diable*, I' *esperaràs* en aquest
 and you.SG devil him.ACC wait.FUT.2SG in this
 loc entrò que El vinga (Vides: 119)
 place until he come.SUBJ.3SG
 'And you, devil, will wait for him in this place until he comes'
- b. M' *esforçaré* al meu poder per serveys
 REFL make an effort.FUT.1SG in.the my power by services
 (Tirant: 1371)
 'I will make the biggest possible effort by services'
- c. Lo acòlith encendrà [...], servirà [...], portarà [...],
 the acolyte light.FUT.3SG serve.FUT.3SG carry.FUT.3SG
 ministrerà se pararà ab camís,
 [...], REFL prepare.FUT.3SG with shirt
 manage.FUT.3SG
 amit y cordó (Baccallar: 267)
 amice and lace
 'The acolyte will light [...], will serve [...], will carry [...], will manage [...], will prepare himself with a shirt, an amice and lace'

Of these three instances provided as examples, the first (5a) can be explained for pragmatic reasons: the vocative, *diable* ('devil'), is preceded by an emphatic subject, *tu* ('you'), signifying that the anteposition of the clitic could be triggered by emphasis (Torres-Latorre 2023, Section 3.1). This pragmatic cause explains the high frequency of SFC-a after a vocative, since all the tokens in this environment have another element before the vocative which could be subject to emphasis. Furthermore, (5a) originates from the *Vides de sants rosselloneses*, a text which has been observed to be highly innovative as regards clitic placement (Torres-Latorre 2023). The reason for this innovation appears to be its diatopic adscription: *Vides de sants rosselloneses* belongs to septentrional Catalan, the Catalan dialect closer to Occitan, and Old Occitan may be more advanced than Catalan as regards both the grammaticalization process of the FC and the fixation of clitic placement in the preverbal position. This hypothesis should be verified in future research. Whatever the case may be, the instances of SFC-a in the first position (5b) and in a paratactic clause (5c) can be understood only as innovative examples of anteposition.

4.2. The Evolution of Clitic Placement in Group 3

The results regarding all environments presented in Table 7 include subordinate clauses and Group 1, that is, the syntactic contexts in which only SFC-a is possible. The difference between preverbal clitics and others, therefore, can be better observed if only the data from Group 3 are taken into account, since this group is made up of those environments in which variation in position is possible (Section 1.1). The quantitative review of Group 3

in Table 9 is more reliable as regards the expansion of anteposition than the general results seen in Table 7.

Table 9. The position of personal pronouns with the FC in Group 3 (11th–16th c.).

| Chronology | SFC-a | AFC or SFC-p |
|------------------------------------|------------------|------------------|
| Preliterary Catalan (11th–13th c.) | 95.8% (23/24) | 4.2% (1/24) |
| 13th century, 2nd half | 41.8% (166/397) | 58.2% (231/397) |
| 14th century, 2nd half | 74.6% (191/256) | 25.4% (65/256) |
| 15th century, 2nd half | 84.8% (179/211) | 15.2% (32/211) |
| 16th century, 2nd half | 96.3% (129/134) | 3.7% (5/134) |
| Total | 67.3% (688/1022) | 32.7% (334/1022) |

However, this reasoning is not yet conclusive, since the variation within Group 3 can sometimes be explained by a more accurate classification in each of the syntactic–pragmatic environments. For instance, emphasis can cause the appearance of a preverbal clitic, as mentioned briefly in Section 1.1. If this is the case, it is necessary to examine the instances of SFC-a one by one in order to determine whether or not they are innovative examples of anteposition. A review of each of the tokens belonging to Group 3 is, therefore, necessary. Only some examples of this group will be commented on in this paper owing to space constraints (see Torres-Latorre n.d.b; cf. Bouzouita and Sentí 2022 and Sentí and Bouzouita 2022 for a tentative study on the effect of pragmatics in these contexts in the 13th and 14th c.), and the focus will be solely on the quantitative view of this phenomenon. Table 9, therefore, provides a summary of the evolution of clitic placement in this group of variation.

The results obtained for the group of variation obey the same development as the total of tokens included in the other groups (Table 7). Preliterary Catalan again differs from the other periods, with SFC-a being attested with by far the most frequency (95.8%, 23/24). The evolution from the 13th c. to the 16th c. shows a noticeable increase in SFC-a: it starts with 41.8% (166/397) in the 13th c. and concludes with 96.3% (129/134) in the 16th c. Indeed, the differences between the four centuries are statistically significant ($\chi^2 = 201.6922$, $p < 0.00001$; significant at $p < 0.05$). Group 3 clearly shows the evolution towards anteposition and the fixation of clitics in this position, regardless of the syntactic context or pragmatic motivations.

It is first necessary to explore the case of preliterary Catalan. This period has the highest level of SFC-a in both the group of variation (see Table 9) and all the possible environments (see Table 7). However, all the preverbal clitics are to be expected, since emphasis can explain anteposition in all the instances, as can be seen in the following example (6), in which the adverb of manner *sí* (‘yes’, ‘this way’) triggers anteposition, as usually occurs with adverbs of manner, not only in Old Spanish (Granberg 1988, pp. 155–94), but also in Old Catalan (Bouzouita and Sentí 2022).

- (6) Sicut superius és escrit sí
as above write.PASS.3SG this way
t' ó tenrei (*Jurament feudal* 3: 71)
you.DAT it.ACC have.FUT.1SG
‘As above it is mentioned, I will have it this way to you’

The gradual increase in SFC-a in the remaining centuries is linked to an increase in the possible environments in which SFC-a can be found. This was the case of Group 2 (Section 4.1) and also of Group 3. Merely as an illustration of this increase, four examples of innovative cases of anteposition are discussed, three after an adverbial phrase (7a–c) and the last after a subordinate clause or an absolute clause (7d). According to the literature on Old Castilian (Granberg 1988, pp. 155–94), adverbs which express a cause and temporal adverbs of chronological progression are followed by postverbal clitics. Nevertheless, from the 14th c. onwards, cases of anteposition are attested in Old Catalan after a causal adverbial phrase (7a) and anteposition is already found after a temporal adverb of chronological progression

in the 13th c. (7b), although it appears in only one text, which is highly innovative for dialectal reasons, as mentioned previously (Torres-Latorre 2023). Examples of preverbal clitics after these temporal adverbs are once again found in the 15th c. (7c). With regard to innovation after subordinate and absolute clauses, the first examples of anteposition are found after an absolute clause in Old Catalan in the 15th c. (7d).

- (7) a. et Déus, per la sua bonea, vos ó
 and God because-of his goodness you.DAT it.ACC
aministrarà
 provide.FUT.3SG
 'and God, because of his goodness, will provide it to you'
 (Muntaner, *Crònica*: 14v)
- b. Eu derocaré aquest temple, e après lo
 I demolish.FUT.1SG this temple and then it.ACC
redificaré
 rebuild.FUT.1SG
 'I will demolish this temple, and then I will rebuild it'
 (Vides: 373)
- c. Après vos farà tres figues e les
 Then you.DAT do.FUT.3SG three figs and them.ACC
 vos posarà en la barba. Après vos
 you.DAT put.FUT.3SG in the beard. Then you.DAT
dirà
 say.FUT.3SG
 'Then he will flip you off in your face. Then he will say'
 (Tirant: 848)
- d. E fet lo dit anniversari, ne
 And hold.PART the mentioned anniversary PART
romandrà
 remain.FUT.3SG in the Cathedral the 10th
 'And held the mentioned anniversary, it will remain in the Cathedral the 10th'
 (Solemnitats: 357)

These four examples illustrate the expansion of SFC-a to environments in which the AFC or SFC-p were previously the usual constructions. A thorough analysis of all of the occurrences of the environments in Group 3 is still necessary in order to review the role played by emphasis in the selection of a preverbal or postverbal clitic position (Torres-Latorre n.d.b). Nevertheless, the data shown in this section confirm the expected evolution towards anteposition in both quantitative and qualitative terms.

5. Conclusions

The study of the grammaticalization of the FC in Old Catalan contributes to understanding the formation of the FC in Ibero-Romance languages. Unlike Old Spanish, in which the SFC with postverbal clitics are rare, the empirical findings in our Catalan corpus attest SFC-p from an early time in the 11th c. and an important token frequency in the 13th and 14th c. (38–39% of the tokens in those syntactic environments that prevent preverbal clitics). The corpus-based data presented in this paper show a more advanced stage of the degree of grammaticalization of the FC in the 13th and the 14th c. in Old Catalan than in other peninsular languages. SFC-p appears frequently in Old Catalan and less use is made of the AFC when compared with Old Navarro-Aragonese and Old Castilian, as has been shown quantitatively. Although some morphological differences among the languages (syncope, rhizotonic verbs) have been suggested in order to partially justify this contrast, the high frequency of the unverbated FC in Old Catalan, even in examples without those morphological features, should be highlighted, since it indicates a more grammaticalized FC. As suggested by Bouzouita (2016), diatopic variation in Old Ibero-Romance could explain the differences in the distribution of the AFC and SFC-p in these languages. The results obtained for Old Catalan appear to confirm a language contact phenomenon in which Catalan (and probably Occitan) spread the more grammaticalized FC (i.e., SFC-p) to Navarro-Aragonese and Castilian (see Primerano et al. 2023).

However, our corpus is unable to confirm the expected increase in frequency of the more grammaticalized construction, i.e., SFC-p in Old Catalan. In fact, the distribution of the AFC and SFC-p with personal pronouns is stable from the 13th c. to the 16th c. This could be accounted for by another ongoing linguistic change process in

diachrony that has been addressed in this paper: clitic positioning. In the 13th–14th centuries, syntactic–pragmatic restrictions governed clitic positioning, and clitics were able to occupy various positions in Old Catalan, unlike in modern and contemporary Catalan (Sentí and Bouzouita 2022). However, this syntagmatic variability disappears in the 16th c., and clitics occupy a fixed slot, mainly a preverbal position, according to the morphological features. This process, which Lehmann (1985) terms as fixation, is part of the grammaticalization of clitic pronouns. This paper outlines the general evolution, although this process should be addressed in more detail in order to, for example, understand the loss of pragmatic conditioning (Torres-Latorre n.d.b). Our empirical corpus data make it possible to draw several conclusions. The first innovative examples with preverbal clitics are attested in the 13th and 14th centuries (preverbal clitics following a vocative), which show the beginning of a change; more examples of preverbal clitics in main sentences (Group 2) are found in the 15th c. (preverbal clitics in the first position of a sentence). It appears that the fixation is almost accomplished in the 16th c., despite some examples of the AFC or SFC-p. This process explains why the SFC-p does not increase in frequency in the Middle Ages. In fact, both the AFC and SFC-p decrease in frequency from the 13th c. onwards in favour of preverbal clitics (SFC-a). The grammaticalization and coalescence of the FC become mixed with another simultaneous grammaticalization process in Old Catalan: that of clitics and the fixation of clitic placement.

Author Contributions: Conceptualization, A.T.-L. and A.S.; methodology, A.T.-L. and A.S.; validation, A.T.-L. and A.S.; formal analysis, A.T.-L.; investigation, A.T.-L. and A.S.; resources, A.T.-L. and A.S.; data curation, A.T.-L. and A.S.; writing—original draft preparation, A.T.-L. and A.S.; writing—review and editing, A.T.-L. and A.S.; visualization, A.T.-L.; supervision, A.S.; project administration, A.T.-L. and A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union, Next Generation EU with a Grant for the Requalification of the Spanish University System (Ministry of Universities of the Government of Spain) (Ref: UV-EXPSOLP2U-1801406) and by the Universitat de València (“Ajudes per a accions especials d’investigació 2022”, Ref. UV-INV_AE-2653948).

Data Availability Statement: The data presented in this study are available in the corpus (CICA 2006) and in some other Old Catalan published texts. The analyzed data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Notes

- ¹ This work has been supported by the research group Vacàlic+ (Universitat de València, Ref. GIUV2013–137).
- ² We shall refer to the “infinitive”, but please note that this element of the Romance construction of the FC does not always coincide with the full form of the infinitive, although diachronically speaking it originates from this verb form. We shall similarly refer to the “auxiliary”, despite the fact that it is no longer a proper auxiliary. See Alsina (2022) for an analysis of the AFC as a compound word consisting of an infinitive, a clitic cluster, and a bound auxiliary.
- ³ In this paper, the term *univerbation* is used as suggested by Lehmann (2020, p. 205), i.e., “the syntagmatic condensation of a sequence of words recurrent in discourse into one word”. It is, moreover, considered to be a gradual process that can have weaker and stronger phases.
- ⁴ There is no agreement about the linguistic status of the AFC. Some authors consider the AFC a periphrasis (Company 2006; Batllori 2015; Octavio de Toledo 2015), but others consider this construction a compound word (Alsina 2022). See Octavio de Toledo (2015, Section 1) and Primerano (2019, Section 3.2.1) for a comprehensive state of the art.
- ⁵ All verb tenses generally require proclisis. However, verbal periphrases admit both proclisis to the finite verb and enclisis to the non-finite one. When the verb appears in the imperative, infinitive or gerund, only enclisis is possible.
- ⁶ There are different approaches to informative structure and different associated terminology (cf. Leonetti and Escandell-Vidal 2021). In this paper, we consider *focus* as the information that is highlighted in a given sentence (see Bouzouita and Sitaridou 2023, Section 2).

For sake of terminological clarity, it should be borne in mind that *focus* is not (always) linked to new information, even though it is quite common.

- 7 This period includes preliterary Catalan (from 11th c. to 1250), the first texts written completely in Catalan (from the 13th c.), medieval Catalan (until 15th c.) and the early modern period (16th c.). The period from the 11th c. to the 16th c. covered is considered to be relatively homogeneous and bounded to the standard Catalan disseminated by the royal *scripta* in the Crown of Aragon.
- 8 The selected texts are the following: for preliterary Catalan (1000–1250): all texts available at CICA (2006) were examined because the extension of texts in this period is, unfortunately, limited; for the 13th c.: *Vides de sants rosselloneses*, *Costums de Tortosa*, *Clams i crims a la València medieval* (1), *Llibre de Cort de Justícia de Cocentaina* (1), *Llibre de Cort de Justícia de Cocentaina* (2), *Crònica de Bernat Desclot* and *Llibre dels Fets*; for the 14th c.: *Regiment de preservació de la pestilència*, *Crònica de Ramon Muntaner*, *Ordinacions con los reys e reynas d’Aragó*, *Epistolari de la València medieval* (I-3 and I-4), *Un llibre reial mallorquí del segle XIV*, *Procés criminal contra Antònia Marquès*, *Manual de consells*, *El procés de Sueca*; for the 15th c.: *Faula de Neptuno i Diana*, *Rahonament fingit entre Francesch Alegre i Sperança*, *Somni de Francesc Alegre*, *Història de les amors de París e Viana*, *Tragèdia de Lançalot*, *Llibre de les solemnitats de Barcelona* (2 and 3), *Lletres reials a la ciutat de Girona* (II-5 and II-6), *Tirant lo Blanch*, *Manual de consells de Gandia a la fi del segle XV* (1); for the 16th c.: *Biografia de Català de Valeriola*, *Constitucions del convent de Sant Josep*, *Diari de Frederic Despalau*, *Edicte del virrei Lluís Vich i Manrique*, *El Llibre de Antiquitats de la Seu de València* (2), *El sínodo del bisbe Baccallar*, *Història general del Regne de Mallorca*, *Instrucció y doctrina que ensenye lo que deu considerar y contemplar lo cristià*, *Les memòries del cavaller valencià Gaspar Antist*, *Llibre del Mostassaf de Mallorca* (14), *Llibre de les Solemnitats de Barcelona* (6), *Los col·loquis de la insigne ciutat de Tortosa*, *Memòries de Jeroni de Saconomina*, *Memòries de Perot de Vilanova*, *Relació de l’entrada i estança a València del rei don Felip II*, *Relació del furt del Santíssim Sacrament a Alcoi*, *Suplicació sobre la cisterna de l’estudi de gramàtica de Randa*. Two texts not included in the CICA (2006) have also been incorporated into our corpus: *Llibre de contemplació en Déu* (by Ramon Llull; 13th c.) and the Gospels of Matthew and John (from the Catalan Bible known as the *Bíblia del segle XIV*, 14th c.; see Bouzouita and Sentí 2022). Some texts have not been studied in their entirety in order to attain an equilibrium between centuries.
- 9 A large quantity of the available texts originate from eastern Catalan in the 14th c., whereas most of them belong to western Catalan in the 15th c., as can be seen in Table 2. This is a well-known difficulty in the study of the history of Catalan.
- 10 Only two tokens are found in preliterary Catalan owing to certain methodological issues. On the one hand, the FC are only two verbal tenses of the entire verbal conjugation of Catalan. Moreover, they are not the most frequent in written texts of the typologies included in our corpus, in which the majority are the present and past tenses. In addition, only the FC with clitic pronouns are studied. According to our empirical findings, SFC-a is the most common construction among the three possible positions of clitics, since it is the only possible construction in subordinate clauses and in Group 1. The AFC and SFC-p, therefore, continue to be limited constructions in quantitative terms. Furthermore, specific reasons can explain the lack of data in certain centuries. As mentioned in the description of the corpus (Section 2), all the texts from CICA (2006) for this period were analysed because of their limited extension. Moreover, the earliest documents usually combine the use of Catalan with the use of Latin, signifying that instances of Catalan FC are scarce.
- 11 We are aware, as one of the reviewers points out, that the stop of the rise in frequency of a grammaticalization construction does not necessarily imply that the construction slows down within the grammaticalization process, and it does definitely not mean that it enters a degrammaticalization process (Bybee 2011, p. 77). Although we do not state that the grammaticalization of FC stops, we deem important the standstill of the relative frequencies of SFC-p and AFC in relation with one another, since not only the more grammaticalized expression does not rise in frequency, but also the less grammaticalized does not seem to disappear if only these two constructions are taken into account. Moreover, the quantitative evolution of SFC-p and AFC in Old Catalan diverges from the evolution of these two forms in other Ibero-Romance languages (see Primerano et al. 2022).
- 12 The data regarding Navarro-Aragonese are taken from Primerano and Bouzouita (2023).
- 13 The data regarding Old Castilian are taken from Primerano et al. (2023).
- 14 See Moreno Bernal (2004) for a morphological study on Spanish future tense focused on the realization of syncope.
- 15 Interestingly, the preference for the AFC with non-adverbial pronouns in this century was more marked than in the total of instances, with 80% of the occurrences in the century (see Table 3).

References

- Alsina, Àlex. 2022. The morphological and syntactic status of the analytic and synthetic future in medieval Catalan. *Catalan Journal of Linguistics* 21: 187–208. [CrossRef]
- Álvarez Blanco, Rosario, Xosé Luís Regueira Fernández, and Xosé Henrique Monteagudo Romero. 1986. *Gramática galega*. Vigo: Galaxia.
- Badia i Margarit, Antoni Maria. 1947. *Los complementos pronominalo-adverbiales derivados de “ibi” e “inde” en la Península Ibérica*. Madrid: CSIC, *Revista de Filología Española*.
- Badia i Margarit, Antoni Maria. 1984. *Gramàtica històrica catalana*. València: Tres i Quatre.
- Batlloori, Montserrat. 2015. Análisis comparativo de las construcciones de dislocación a la derecha y su incidencia en el orden de palabras del español, catalán y portugués medievales. In *El orden de palabras en la historia del español y otras lenguas iberorromances*. Edited by Mónica Castillo Lluich and Marta López Izquierdo. Madrid: Visor, pp. 83–110.

- Batllo, Montserrat, and M. Lluïsa Hernanz. 2015. Weak Focus and Polarity: Asymmetries between Spanish and Catalan. In *Syntax over Time. Lexical, Morphological and Information-Structural Interactions*. Edited by Theresa Biberauer and George Walkden. Oxford: Oxford University Press, pp. 280–98.
- Batllo, Montserrat, Narcís Iglésias, and Ana Maria Martins. 2005. Sintaxi dels clítics pronominals en el català medieval. *Caplletra* 38: 137–77.
- Bonet, Eulàlia. 2002. Cliticització. In *Gramàtica del català contemporani*. Edited by Joan Solà, Maria-Rosa Lloret, Joan Mascaró and Manuel Pérez Saldanya. Barcelona: Empúries, vol. 1, pp. 933–89.
- Bouzouita, Miriam, and Andreu Sentí. 2022. La gramaticalización del futuro y el condicional en el iberorromance del siglo XIV a partir de traducciones bíblicas paralelas: El caso del castellano y el catalán antiguos. In *Traducción bíblica e historia de las Lenguas Iberorrománicas*. Edited by Andrés Enrique-Arias. Berlin: W. de Gruyter, pp. 243–73.
- Bouzouita, Miriam, and Ioanna Sitaridou. 2023. Cambios en el orden de palabras y la estructura informativa. In *Linguística histórica del español/The Routledge Handbook of Spanish Historical Linguistics*. Edited by Steven Dworkin, Gloria Clavería and Álvaro Octavio de Toledo y Huerta. London: Routledge, pp. 221–32, in press.
- Bouzouita, Miriam. 2008a. The Diachronic Development of Clitics in the History of Spanish. Ph.D. dissertation, King's College, London, UK.
- Bouzouita, Miriam. 2008b. At the Syntax-Pragmatics Interface: Clitics in the History of Spanish. In *Language in Flux: Dialogue Coordination, Language Variation, Change and Evolution*. Edited by Robin Cooper and Ruth Kempson. London: College Publications, pp. 221–63.
- Bouzouita, Miriam. 2011. Future Constructions in Medieval Spanish: Mesoclisism Uncovered. In *The Dynamics of Lexical Interfaces*. Edited by Ruth Kempson, Eleni Gregoromichelaki and Chris Howes. Stanford: CSLI Publications, pp. 91–132.
- Bouzouita, Miriam. 2012. Los futuros analíticos y sintéticos en la Fazienda de Ultra Mar. In *Actas del VIII Congreso Internacional de Historia de la Lengua Española*. Edited by Enrique Montero Cartelle. Santiago de Compostela: Meubooks, vol. 2, pp. 1631–42.
- Bouzouita, Miriam. 2013. La influencia latinizante en el uso del future en la traducción bíblica del código Escorial I.i.6. In *Actes del 26é Congrès de Lingüística i Filologia Romàniques (València, 6–11 de setembre de 2010)*. Edited by Emili Casanova Ferrero and Cesáreo Calvo Rigual. Berlin: W. de Gruyter, pp. 353–64.
- Bouzouita, Miriam. 2016. La posposición pronominal con futuros y condicionales en el código escorialense I.i.6: Un examen de varias hipótesis morfosintácticas. In *Linguística de Corpus y Lingüística Histórica Iberorrománica*. Edited by Johannes Kabatek. Berlin: W. de Gruyter, pp. 271–301.
- Buridant, Claude. 2019. *Grammaire du français médiéval (XI^e–XIV^e siècles)*. Strasbourg: Société de Linguistique Romane, Éditions de linguistique et philologie.
- Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In *The Handbook of Historical Linguistics*. Edited by Brian D. Joseph and Richard D. Janda. Oxford: Blackwell, pp. 602–23.
- Bybee, Joan. 2011. Usage-based theory and grammaticalization. In *The Oxford Handbook of Grammaticalization*. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press, pp. 69–78.
- Castillo Lluch, Mónica. 1996. La posición del pronombre átono en la prosa hispánica medieval. Ph.D. dissertation, Universidad Autónoma de Madrid, Madrid, Spain.
- Castillo Lluch, Mónica. 2002. Distribución de las formas sintéticas y analíticas de futuro y condicional en español medieval. In *Actas del V Congreso Internacional de Historia de la Lengua Española*. Edited by María Teresa Echenique Elizondo and Juan Pedro Sánchez Méndez. Madrid: Gredos, pp. 541–50.
- CICA. 2006. Corpus Informatitzat del Català Antic. Directed by Joan Torruella, Manuel Pérez Saldanya and Josep Martines. Available online: <http://www.cica.cat> (accessed on 15 January 2023).
- Clavería, Gloria, and Joan Torruella. 2012. El Corpus informatizado del catalán antiguo (CICA) y su herramienta de explotación (Estación de análisis documentales, EAD). In *Nuevas Perspectivas Para la edición y el Estudio de Documentos Hispánicos Antiguos*. Edited by María Jesús Torrens Álvarez and Pedro Sánchez-Prieto Borja. Berna: Peter Lang, pp. 405–25.
- Coleman, Robert. 1971. The Origin and Development of Latin Habeo+Infinitive. *The Classical Quarterly* 21: 215–32. [CrossRef]
- Company, Concepción. 1985. Los futuros en el español medieval, sus orígenes y su evolución. *Nueva revista de filología hispánica* 34: 48–107. [CrossRef]
- Company, Concepción. 2006. Tiempos de formación romance II: Los futuros y condicionales. In *Sintaxis histórica de la lengua española*. Edited by Concepción Company. México: Fondo de Cultura Económica/Universidad Nacional Autónoma de México, vol. 1, pp. 347–418.
- Company, Concepción. 2012. Historical morphosyntax and grammaticalization. In *The Handbook of Hispanic Linguistics*. Edited by José Ignacio Hualde, Erin O'Rourke and Antxon Olarrea. Londres/New York: Blackwell, pp. 673–93.
- Company, Concepción, and Alfonso Medina. 1999. Sintaxis motivada pragmáticamente: Futuros analíticos y futuros sintéticos en el español medieval. *Revista de Filología Española* 79: 65–100. [CrossRef]
- Cunha, Celso, and Luís F. Lindley Cintra. 1992. *Gramática do Português Contemporâneo*. Lisboa: Edições João Sá da Costa.
- de Andrés Díaz, Ramón. 2013. *Gramática comparada de las lenguas ibéricas*. Gijón: Trea.
- Eberenz, Rolf. 1991. Futuro analítico y futuro sintético en tres obras con rasgos coloquiales: El 'Corbacho', 'La Celestina' y 'La Lozana Andaluza'. In *Homenaje a Hans Flasche: Festschrift zum 80. Geburtstag 25. November 1991*. Edited by Rafael Lapesa Melgar. Stuttgart: Franz Steiner Verlag, pp. 499–508.

- Elvira, Javier. 2015. ¿V2 en español antiguo? In *El orden de palabras en la historia del español y otras lenguas iberorromances*. Edited by Mónica Castillo Lluch and Marta López Izquierdo. Madrid: Visor, pp. 27–47.
- Fischer, Susann. 2002. *The Catalan Clitic System. A Diachronic Perspective on Its Syntax and Phonology*. Berlin: W. de Gruyter.
- Fischer, Susann. 2005. Construcciones con avance estilístico en el catalán antiguo. *Caplletra* 38: 119–35.
- Fleishman, Susanne. 1982. *The Future in Thought and Language. Diachronic Evidence from Romance*. Cambridge: Cambridge University Press.
- GIEC Institut d'Estudis Catalans (IEC). 2016. *Gramàtica de la Llengua Catalana*. Barcelona: IEC.
- Granberg, Robert Arthur. 1988. Object Pronoun Position in Medieval and Early Modern Spanish. Ph.D. dissertation, University of California, Los Angeles, CA, USA.
- Granberg, Robert Arthur. 1999. Clitic Position in Thirteenth-Century Spanish: Sentences with Preverbal Subjects. *La Corónica* 27: 89–113.
- Hopper, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Jensen, Frede. 1994. *Syntaxe de l'ancien occitan*. Tübingen: Niemeyer.
- Kabatek, Johannes. 2005. Tradiciones discursivas y cambio lingüístico. *Lexis: Revista de lingüística y literatura* 29: 151–77. [CrossRef]
- Kabatek, Johannes. 2013. ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania: Revista dedicada a las lenguas y literaturas iberorrománicas de Europa y América* 77: 8–28. [CrossRef]
- Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20: 303–18.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization*, 2nd ed. Edition of *Thoughts on Grammaticalization: A Programmatic Sketch*, 1982. Munich: Lincom Europa.
- Lehmann, Christian. 2020. Univerbation. *Folia Linguistica Historica* 41: 205–52. [CrossRef]
- Leonetti, Manuel, and M. Victoria Escandell-Vidal. 2021. *La estructura informativa*. Madrid: Visor.
- Martines, Josep, and Manuel Pérez Saldanya. 2009. Per a una gramàtica del Català Antic: Objectius, metodologia i un tast. In *Per a una gramàtica del català antic*. Edited by Manuel Pérez Saldanya and Josep Martines. Valencia: Institut Interuniversitari de Filologia Valenciana, pp. 7–39.
- Martins, Ana Maria. 1994. Clíticos na história do português. Ph.D. dissertation, University of Lisbon, Lisbon, Portugal.
- Mathieu, Eric. 2006. Stylistic fronting in Old French. *Probus* 18: 219–66. [CrossRef]
- Matute, Cristina. 2016. Entre pronombres y adverbios: Mecanismo de cambio en la historia dialectal peninsular de *hi/j* < *ibi*. *Boletín de la Real Academia Española* 96: 201–37.
- Meilán García, Antonio José. 1994. Funcionamiento y valores del pronombre *ende* en el castellano antiguo. *Revista de Filología de la Universidad de La Laguna* 13: 245–62.
- Moll, Francesc de B. 2006. *Gramàtica històrica catalana*. Valencia: Publicacions de la Universitat de València. First published 1952.
- Moreno Bernal, Jesús. 2004. La morfología de los futuros románicos. Las formas con metátesis. *Revista de filología románica* 21: 121–69.
- Octavio de Toledo, Álvaro. 2015. Futuros que se miran el ombligo: Mesoclisys y anteposición de formas no personales en la historia del español. In *El orden de palabras en la historia del español y otras lenguas iberorromances*. Edited by Mónica Castillo Lluch and Marta López Izquierdo. Madrid: Visor, pp. 141–234.
- Oliva, Salvador, and Pep Serra. 2002. Accent. In *Gramàtica del català Contemporani*. Edited by Joan Solà, Maria-Rosa Lloret, Joan Mascaró and Manuel Pérez Saldanya. Barcelona: Empúries, vol. 1, pp. 345–91.
- Paden, William Doremus. 1998. *An introduction to Old Occitan*. New York: The Modern Language Association of America.
- Pérez Saldanya, Manuel. 1998. *Del llati al català. Morfosintaxi verbal històrica*. Valencia: Publicacions de la Universitat de València.
- Polo Cano, Nuria. 2006. Breve aproximación a los adverbios pronominales “y” y “ende” en documentos notariales de la Vizcaya tardomedieval. *Oihenart* 21: 377–93.
- Primerano, Antoine. 2019. La gramaticalización de los futuros y condicionales en el navarroaragonés de los siglos XIII y XIV: Un análisis morfosintáctico-pragmático. Master dissertation, University of Ghent, Ghent, Belgium.
- Primerano, Antoine. 2021. Los derivados de *ibi* e *inde* en navarroaragonés antiguo: Una aproximación a su sintaxis. In *Quand le syntagme nominal prend ses marques: Du prédicat à l'argument*. Edited by Peter Lauwers, Katia Paykin, Mihaela Illoaia, Machteld Meulleman and Pascale Hadermann. Reims: Épure, pp. 405–25.
- Primerano, Antoine, and Miriam Bouzouita. 2023. La gramaticalización de los futuros y condicionales en el navarroaragonés de los siglos XIII y XV: Una comparación con el castellano medieval. *Revue Romane*, in press.
- Primerano, Antoine, Aina Torres-Latorre, Andreu Sentí, and Miriam Bouzouita. 2022. Morfosintaxis y pragmática del futuro y condicional de las lenguas iberorrománicas. Paper presented at XXX Congreso internacional de lingüística y filología románicas, La Laguna, Spain, July 6.
- Primerano, Antoine, Aina Torres-Latorre, Andreu Sentí, and Miriam Bouzouita. 2023. La gramaticalización del futuro y el condicional en iberorromance antiguo: Castellano y catalán en contacto. *Revista de historia de la lengua española*, in press.
- Ribera, Josep Enric. 2020. *La sintaxi dels pronoms clítics no personals en català antic*. Alacant: Universitat d'Alacant/Departament de Filologia Catalana.
- Saralegui, Carmen. 1985. Sobre la morfología de algunos futuros y condicionales de las hablas hispánicas. *Homenaje a Álvaro Galmés de Fuentes* 2: 203–12.
- Sentí, Andreu, and Miriam Bouzouita. 2022. The Syntax of Old Catalan Clitics: Llibre dels Fets. *Catalan Journal of Linguistics* 21: 47–77. [CrossRef]

- Torres-Latorre, Aina. 2022. La gramaticalització del futur i el condicional en català antic (segles XIII-XV): Un estudi de corpus. Master's dissertation, Universitat Autònoma de Barcelona/Universitat de Barcelona, Barcelona, Spain.
- Torres-Latorre, Aina. n.d.a. Formes analítiques o formes sintètiques: Factors morfològics que condicionen la dualitat formal del futur i el condicional en català antic, in preparation.
- Torres-Latorre, Aina. n.d.b. Productivitat i pragmàtica en la sintaxi dels clítics pronominals en català antic. *Anuari de Filologia. Estudis de lingüística*, submitted.
- Torres-Latorre, Aina. 2023. La gramaticalització del futur i el condicional al català del segle XIII: Formes analítiques i formes sintètiques. *Zeitschrift für Katalanistik* 26, in press.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Not Only Anteriority in the Past: The Functions of the Pluperfect in Spoken Italian

Eleonora Morei

Faculty of Education, Free University of Bozen-Bolzano, Viale Ratisbona 16, 39042 Bressanone, Italy; emorei@unibz.it

Abstract: Three distinct anaphoric functions and one deictic function are, with fair confidence, associated with the Italian Pluperfect in the existing literature. In recent studies, it has been hypothesized that the Italian Pluperfect may also have an aoristic use. The present study attempts to assess the semantics of the Italian Pluperfect, by a corpus-based methodology. It will be shown that the data do not support the hypothesis of an aoristic use of the Pluperfect: rather, they suggest the need to extend the analysis of the Pluperfect's semantics to domains other than tense and aspect. It will be argued that (inter)subjectification may have a key role in describing the layered semantics of the Italian Pluperfect, especially concerning its possible modal-evidential developments.

Keywords: Pluperfect; discontinuous past; Italian

1. Introduction

The Italian Pluperfect displays a rather prototypical semantic core, with four distinct temporal–aspectual functions that have been identified by previous research: past-in-the-past; perfect-in-the-past; reversed result; and past temporal frame (Squartini 1999). In recent studies (Bertinetto 2003, 2014; Bertinetto and Squartini 1996; Scarpel 2017), it has also been hypothesized, albeit not specifically dealt with, that the Pluperfect may have an additional aoristic use in spoken Northern Italian. This paper aimed to assess the existence of such a use by analyzing Pluperfect occurrences in ParlaTO (Cerruti and Ballarè 2020), a corpus of spontaneous speech collected in Turin between 2018 and 2020. It will be shown that the data did not support the hypothesis of an aoristic use, but that they suggested that the Italian Pluperfect was developing secondary semantic values that could be explained by taking (inter)subjectification¹ paths Traugott (2003, 2010) of grammaticalization into account.

The relationship between grammaticalization and (inter)subjectification has been extensively discussed by Traugott (2010). Having once defined grammaticalization as “[t]he change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and once grammaticalized, continue to develop new grammatical functions” (Hopper and Traugott 2003), Traugott (2010) states that subjectification is likely to occur in grammaticalization “presumably because grammaticalization by definition involves recruitment of items to mark the speaker’s perspective on [a series of] factors”, among which are tense (“how the proposition (ideational expression) is related to speech time or to the temporality of another proposition”) and aspect (“whether the situation is perspectivized as continuing or not”), i.e., the categories that the present study is mostly concerned with, but also modality (“whether the situation is relativized to the speaker’s beliefs”) and discourse markers (“how utterances are connected to each other”), i.e., those categories that are known to be mostly involved in the Pluperfect’s development of secondary meanings (Plungian and van der Auwera 2006).

In what follows, it will be shown how tense, aspect and modality intertwine in defining the Pluperfect’s semantics. Concerning modality, it will be argued in Sections 3 and 4 that possible Pluperfect developments include modal-evidential values, for the analysis of which the interplay of subjective and intersubjective values may be especially relevant.

Citation: Morei, Eleonora. 2023. Not Only Anteriority in the Past: The Functions of the Pluperfect in Spoken Italian. *Languages* 8: 82. <https://doi.org/10.3390/languages8010082>

Academic Editors: Marlies Jansegers, Renata Enghels and Tom Bossuyt

Received: 15 January 2023

Revised: 2 March 2023

Accepted: 6 March 2023

Published: 13 March 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

As outlined in the existing literature, the prototypical meaning of the Pluperfect is locating an Event (E) prior to a Reference Time (RT), at which the resulting state of E holds, and which is, in turn, in the past, with respect to the Speech Time (ST), as in:

- (1) Bill had arrived at six o'clock (Comrie 1976)

Nevertheless, as Comrie (1976) explains, Figure 1 accounts for just one of the possible readings of the sentence *Bill had arrived at six o'clock*. It could also be interpreted that Bill arrived precisely at six o'clock, before an unspecified RT.

In the case of Figure 1, the event encoded by the Pluperfect (Bill's arrival) is related to the state of affairs at six o'clock (Bill still being there), and Comrie (1976) therefore dubs this reading as perfect-in-the-past. In the case of Figure 2, the event encoded by the Pluperfect (Bill's arrival) is understood to precede another (unknown) past event, but it is not related to it, and Comrie (1976) therefore dubs this reading as past-in-the-past.

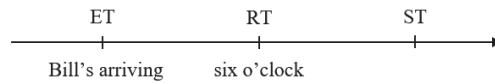


Figure 1. Perfect-in-the-past reading

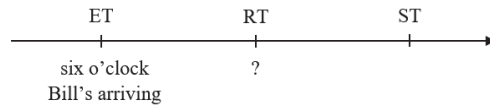


Figure 2. Past-in-the-past reading

Beyond recognizing both these functions as pertaining to the semantics of the Italian Pluperfect, Squartini (1999) identifies two additional functions that the Italian Pluperfect may have, i.e., reversed result and past temporal frame. The reversed result function emphasizes that the result of the event encoded by the Pluperfect is no longer valid.

- (2) Me lo aveva promesso, ma adesso fa finta di non ricordarsene (Squartini 1999)
'(S)He had promised me, but now (s)he acts as if (s)he didn't'

Given that the event encoded by the Pluperfect is related to the (reversed) state of affairs that holds at a later time, the reversed result function can be interpreted as a subcategory of the perfect-in-the-past function (Squartini 1999), as depicted in Figure 3.

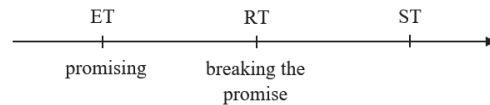


Figure 3. Reversed result reading

On the other hand, the past temporal frame function stresses the idea that the event encoded by the Pluperfect belongs to a closed temporal section that is grounded in the past.

- (3) Su questo argomento tanti anni fa N. ci aveva scritto un libro (Squartini 1999)
'N. wrote a book on this many years ago'

This is the only function, amongst the four functions Squartini (1999) associates with the Italian Pluperfect, that is deictic rather than anaphoric, i.e., its representation does not

involve an RT. This distinguishes the past temporal frame function from the past-in-the-past function, in the sense that, whereas a past-in-the-past Pluperfect places the event in the past with respect to an RT, a past-temporal-frame Pluperfect places the event in the past with respect to the ST directly, with no relationship of anteriority being identifiable (see Figure 4).

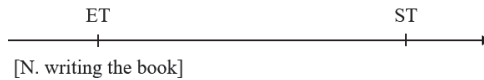


Figure 4. Past temporal frame reading

In recent studies (Bertinetto 2003, 2014; Bertinetto and Squartini 1996; Scarpel 2017), it has been hypothesized that the Pluperfect can have an aoristic (i.e., deictic) use in spoken Northern Italian. It is important to note that while being deictic, the past temporal frame function cannot be described as aoristic, as it does not concern past events in general, but rather a smaller number of cases, i.e., events whose non-relevance at the ST is stressed². On the other hand, an alleged aoristic Pluperfect should be able to potentially encode all cases of perfectivity in the past (except for the Perfect type).

Undoubtedly, the two ‘new’ functions identified by Squartini (1999)—reversed result and past temporal frame—have a strong subjective component, given that they encode the speaker’s perspective on the present (ir)relevance of past events. Furthermore, the past temporal frame function can also be understood as subjective, in that deictic grammatical functions “localize the linguistic entity they apply to with respect to the coordinates of the speaker” (Diewald 2011). It will be shown in Section 4 that additional semantic values arising from that data, while not always closely related one to the other, can all be subsumed under the category of (inter)subjectivity.

2. Materials and Methods

The research was carried out on ParlaTO³ (Cerruti and Ballarè 2020), a corpus of spontaneous speech collected through semi-structured interviews in Turin between 2018 and 2020. ParlaTO is designed to account for diastratic variation in the Italian spoken in Turin: it is balanced for the speaker’s age (16–29, 30–59 and 60+ years old), and provides a large set of additional metadata (Cerruti and Ballarè 2020). However, as the bibliography did not suggest that the aoristic use of the Pluperfect was related to a particular diastratic variety, the corpus was queried in its entirety as being simply representative of a variety of Northern Italian. One additional aspect of ParlaTO to consider when discussing the results of the present study is that the interviewees were often encouraged to talk about the past or to share memories: as a result, the corpus presented a high number of contexts relating to personal past experiences.

As the ParlaTO corpus was not POS tagged, the following query was used to extract Pluperfect occurrences:

```
[word = "avev.*|er.*"] [word = ".*at.?|. *ut.?|. *it.?| fatto| detto| visto| messo| preso"]
```

Given that the Italian Pluperfect is formed with the Imperfect of the auxiliary verb and the Past Participle of the lexical verb, the query searched for the Imperfect of one of the possible auxiliary verbs (*avere* ‘to have’ or *essere* ‘to be’) followed either by a word ending with regular Past Participle desinences or by one of the most frequent irregular Past Participles (*fatto* ‘done’, *detto* ‘said’, *visto* ‘seen’, *messo* ‘put’, *preso* ‘taken’) according to the frequency list of the LIP corpus (De Mauro et al. 1993).

The query, run on 12 April 2021, produced 600 results: given the length of the analysis, it was decided to limit it to the first 300 results, 245 of which were actual Pluperfect occurrences.

3. Results

3.1. Past-in-the-Past

The Pluperfects classified as instances of past-in-the-past numbered 145 (out of a total of 245). The clearest examples displayed adverbial modifiers (underlined in the following example) that clarified the deictic/anaphoric collocation of the events referred to:

- (4) TOI052: e quindi stavo facendo fare
 TOI052: ehm
 TOI052: dei lavoretti per pasqua e pasquetta
 TOI051: si
 TOI052: e invece una volta prima gli *avevo fatto* [113] fare delle cornici
 TOR007: che carini
 TOI052: e dei quadretti cose così' e poi *avevamo fatto* [114] proprio la mostra eh ce'

 TOI052: and so I was having them make
 TOI052: uhm
 TOI052: Easter crafts
 TOI051: yes
 TOI052: and one time before I *had had* [113] them make frames
 TOR007: how cute
 TOI052: and small paintings things like that and then we *had done* [114] a proper exhibit I mean⁴

In (4), TOI052 was speaking about her volunteering experiences by showing pictures. The modifier *una volta prima* 'one time before' leaves no doubt about the fact that Pluperfects [113] and [114] preceded the event spoken about in the first lines (i.e., the making of Easter crafts).

In other cases, the anteriority of the Pluperfect was recognizable only because of world knowledge, by which we knew how certain events usually preceded others:

- (5) TOR004: si' ma sai che *avevo sentito* [22] di uno che *aveva fatto* [23] causa a starbucks perche' non aveva scritto che il bicchiere poteva essere bollente quello si *era bruciato* [24] la mano

 TOR004: yeah but do you know that I heard (lit. *had heard* [22]) about someone who *had sued* [23] Starbucks because they didn't write (lit. hadn't written) that the cup could've been hot that one *had burned* [24] his hand

In (5), the only possible collocation of the events was: person burns his hand > person sues Starbucks > TOR004 hears about this. While [23] and [24] were therefore past-in-the-past Pluperfects, [22] seemed to be deictic, and was classified as a past-temporal-frame Pluperfect. In fact, (5) also contained one additional Pluperfect (*non aveva scritto* 'they hadn't written') that was not italicized in the text and was not associated with a number: this was because, being formed with an irregular past participle, it could not be identified by the query. Although only the Pluperfects identified by the query were included in the counts, it could still be observed that *non aveva scritto* 'they hadn't written' in turn displayed anteriority with respect to [24], and was therefore also classified as a past-in-the-past Pluperfect.

For other cases, the anteriority value of the Pluperfect was unquestionable, yet it was not the most salient, as illustrated in example (6):

- (6) TOI008: san giovanni che io ho trovato incredibile perche' eh mh
 TOI008: appunto io abito vicino al po quindi ci metto cinque minuti ad andare li'
 TOR001: mh mh
 TOI008: e tutti gli anni sempre andato tantissima gente sempre strapieno quest'anno eh mh ho detto non vado nemmeno perche' ci *avevano messo* [64] i tornelli il

TOR001: no certo

TOI008: tutti i vari controlli e ho detto non ci vado nemmeno

TOI008: San Giovanni⁵ which I found incredible because uh uhm

TOI008: indeed I live close to the Po so I'm there in five minutes

TOR001: uhm uhm

TOI008: and every year I always went lots of people always super full this year uh uhm I said I'm not even going because they *had put* [64] the turnstiles the

TOR001: no of course

TOI008: all the controls and I said I'm not even going

Pluperfect [64], for instance, seemed to emphasize the cause/effect relationship between the placement of the turnstiles and the decision not to attend the event, rather than the temporal anteriority of the former with respect to the latter. Nevertheless, in all such cases where an anteriority relationship could still be identified, the Pluperfects under scrutiny were classified as past-in-the-past: this may also explain the numerical prevalence of the Pluperfects categorized as such.

3.2. Perfect-in-the-Past

The Pluperfects classified as instances of perfect-in-the-past numbered 28, which was likely an undervaluation of the actual occurrence of this category, since the query could only identify Pluperfects if no word(s) occurred between the auxiliary and the lexical verb, which meant that Pluperfects combined with the adverb *già* 'already' (highly compatible with a perfect reading)—as in *avevo già mangiato* 'I had already eaten'—were excluded from the analysis. Nevertheless, the potential compatibility of the Pluperfects under scrutiny with the adverb *già* 'already' was an important clue to guiding their classification as perfect-in-the-past, as exemplified in (7).

- (7) TOI065: eravamo frutto di una mh
 TOI065: di una riforma
 TOR001: mh mh
 TOI065: e non potevamo passare al secondo biennio era ancora quadriennale
 TOI065: se non *avevamo dato* [60] quattro obbligatori del primo
 TOR001: okay
 TOI065: ma eh
 TOI065: io ce n'era ne *avevo dati* [61] due gli altri due erano enormi

TOI065: we were the result of a uhm

TOI065: of a reform

TOR001: uhm uhm

TOI065: and we couldn't move to the third and fourth years it was still four years long

TOI065: if we *hadn't passed* [60] four mandatory [exams] of the first two years

TOR001: okay

TOI065: but uh

TOI065: I there was I *had passed* [61] two the other two were huge

It was not only Pluperfects [60] and [61] that were undoubtedly compatible with the adverb *già* 'already'⁶—given that the students needed to have already passed four exams to continue their course of study—but also the resultative value of the Pluperfect, which was evident from the fact that having (not) passed four exams was extremely relevant at the RT (i.e., the start of the second half of the course of study).

In other cases, the compatibility of the Pluperfect with the adverb *già* 'already', seemed to depend on matters of interpretation, as in the following case:

- (8) TOI077: le parole testuali aveva fatto
 TOR004: fai occhio
 TOI077: han fatto effetto evidentemente
 TOR004: quanto te le sei preparate
 TOI077: no niente perche' non e'
 TOI077: non avevo intenzione
 TOR004: non era neache previsto
 TOI076: non voleva
 TOI077: si' *avevo capito* [31] che x vabbe' questa⁷
 TOI077: vuol⁸ che io x che la accompagno
- TOI077: he said those literal words
 TOR004: watch out
 TOI077: clearly they worked
 TOR004: how much did you prepare them
 TOI077: no nothing because it's not
 TOI077: I had no intention
 TOR004: it wasn't even planned
 TOI076: he didn't want
 TOI077: yes I *had understood* [31] that well this one
 TOI077: she wants me to give her a ride

In the part of the conversation preceding (8), TOI076 talked about when TOI077 (who was her boss at the time) finally agreed to give her a ride, and confessed to reciprocating her feelings. As its object was not expressed, Pluperfect [31] could have had two readings: either TOI077 already understood that TOI076 wished to get a ride from him because she had feelings for him (perfect-in-the-past, compatible with *già* 'already'), or TOI077 initially (mis)understood that TOI076 simply had a genuine need for a ride (reversed result)⁹. In such cases, reading the entire conversation was essential to determining which interpretation was the most likely (e.g., the perfect-in-the-past reading for (8), as confirmed by the audio track, given that TOI077 speaks with a mocking tone).

There were also cases where clues others than compatibility with *già* 'already' had to be taken into account:

- (9) TOI051: perche' l l la sua mamma era podalica e' stato un parto bruttissimo
 TOR007: mh anch'io sono nata podalica
 TOI051: eh
 TOI051: mah
 TOR007: eh ha sofferto molto mia mamma
 TOI051: ma io da una parte era solo un un anno e mezzo che *avevo avuto* [118] il primo figlio
 TOI052: mh mh
 TOI051: e allora le ossa erano ancora
 TOI051: eh
 TOR007: si' si'
 TOI051: abbastanza aperte
- TOI051: because her mom was podalic it was a terrible delivery
 TOR007: uhm I was born podalic too
 TOI051: uhm
 TOI051: bah
 TOR007: uhm she suffered a lot my mom
 TOI051: well on one hand for me it had been just a year and a half since I had had [118] my first child
 TOI052: uhm uhm

TOI051: and so the bones were still
 TOI051: uhm
 TOR007: yes yes
 TOI051: pretty open

Pluperfect [118] was not really compatible with *già* 'already'; nevertheless, the expression *era solo un anno e mezzo che* 'it had been just a year and a half since' measured the temporal distance of Pluperfect [118] from the RT, and therefore suggested a perfect reading.

3.3. Reversed Result

The Pluperfects classified as instances of reversed result numbered 16. In many cases, these Pluperfects conveyed exactly the opposite meaning of an Italian Present Perfect¹⁰.

- (10) TOI003: ha perso tantissime cose torino
 TOR001: mh
 TOI003: se uno pensa
 TOI002: no pero' si e' arricchita parecchio con le olimpiadi a pa guarda prima non c'era
 TOI003: a partire da esperimenta
 TOI003: a partire da un macello di cose che io mi ricordo quando andavo a scuola potevi fare
 TOI003: un casi *era diventata* [187] la citta' delle delle mh
 TOI002: mado' ma prima tu vede mado' ma tu prima vedevi turismo a torino
 TOI003: del libro e poi l'ha spostata a milano
- TOI003: Torino lost a lot of things
 TOR001: uhm
 TOI003: if one thinks
 TOI002: no but it developed a lot with the Olympics look before there wasn't
 TOI003: starting from Esperimenta
 TOI003: starting from a lot of things that I remember when I still went to school you could do
 TOI003: a lot it became (lit. *had become* [187]) the city of of uhm
 TOI002: God but before you saw God but before you saw tourism in Torino
 TOI003: of books and then they moved it to Milano

In (10), not only did TOI003 explicitly provide a motivation for the results of the event (i.e., Turin becoming the city of books) being considered as reversed (the city of books was now Milan), but also Pluperfect [187] itself encoded this semantic. Had a Present Perfect replaced Pluperfect [187], it would likely have been inferred that Turin was still the city of books¹¹.

While Pluperfect [187] (and most of the other reversed-result Pluperfects analyzed) displayed exactly the opposite semantic of a Perfect result (see Comrie (1976) for a description of types of Perfect, and Bertinetto (1986) for Italian), there were also instances of reversed-result Pluperfects functioning as the opposite of a Perfect of persistent situation:

- (11) TOI054: e quello mi *aveva fatto* [88] mi aveva un po' pero'
 TOI054: poi mi e' passato
- TOI054: and that did (lit. *had done* [88]) me a little but
 TOI054: then it went away

TOI054 was referring to an alarming road trip that she had experienced, and likely meant to say that it scared her (*mi aveva fatto paura* 'it scared (lit. *had scared*) me'). While in (10) the Pluperfect reversed the result of the event, in (11) the Pluperfect reversed the state of affairs itself, given that it stopped taking place. In other words, Pluperfect [88]'s meaning

could be understood as *it scared me, but it doesn't anymore*, while the meaning of Pluperfect [187] was *Turin became the city of books but it isn't anymore*, rather than **Turin became the city of books, but it doesn't become anymore*. This will be further discussed in Section 4.

In other cases where 'anti-Perfect' semantics were not particularly evident, contextual and cotextual clues played an important role in guiding the classification of the Pluperfects. An example thereof is the following:

- (12) TOI077: ci hanno proposto il viaggio al cairo in pullman abbiam detto vabbe' quando ci ricapita
 TOR004: perche' era organizzato
 TOI077: si'
 TOR004: okay
 TOI076: e c *avevano detto* [44]
 TOI076: meno ore e invece poi alla fine siam stati sei ore in quel
 TOI077: si'
 TOI077: cinque sei ore
 TOI076: sei ore

TOI077: they proposed us the trip to Cairo by bus we said well this won't happen a second time
 TOR004: because it was organized
 TOI077: yes
 TOR004: okay
 TOI076: and they told (lit. *had told* [44]) us
 TOI076: less hours and instead then in the end we've stayed six hours in that
 TOI077: yes
 TOI077: five six hours
 TOI076: six hours

Given that receiving information about the flight's length must have preceded the landing, Pluperfect [44] could be understood as having been a past-in-the-past. Nevertheless, as TOI076 clearly stated that the results of the event encoded Pluperfect [44] had been reversed (the information turned out to be wrong), it was classified as an instance of reversed result.

3.4. Past Temporal Frame

The Pluperfects classified as instances of past temporal frame numbered 63. It was not surprising that many past-temporal-frame Pluperfects occurred in contexts of remembering, and were often signaled as such by the speakers themselves:

- (13) TOR004: eh ma tortoli' eh mizzeca e' bellissima
 TOR004: c'e' la spiaggia del saraceno quello con la torre ti ricordi *eravamo andati* [48] anche insieme l'anno che sei venuta
 TOR004: uh but Tortoli uh my goodness is really beautiful
 TOR004: there is the Saraceno beach that with the tower do you remember we also went (lit. *had gone* [48]) together the year you came

By asking the addressee whether she remembered, TOR004 implicitly assigned the event encoded by Pluperfect [48] to a temporal frame (that of the memory) past and closed, further specified by the temporal indication *l'anno che sei venuta* 'the year you came'. It is worth clarifying that Pluperfect [48] was indeed deictic: even by analyzing a broader section of the conversation, it was not possible to retrieve any RT, and the sentence would also have been acceptable if a Present Perfect (in its aoristic function) had replaced Pluperfect [48] (*siamo andati* 'we went', lit. 'we have gone').

In other cases, the belonging of the event encoded by the Pluperfect to a past temporal frame could be signaled by expressions with the noun *volta* ‘time’ (e.g., *la volta* ‘the time (that)’, *la prima volta* ‘the first time (that)’, etc.) or simply inferable from context:

- (14) TOR004: avevi degli amici nel paese dove vivevi
 TOI054: no io tantissimi amici
 TOI054: sempre avuto tante c tante conoscenze ma tanti amici anche tanta gente che le piaceva stare con me
 TOR004: e ma organizzavate delle feste facevate delle cose
 TOI054: perche’ comunque
 TOI054: si’ anche a casa mio padre per i miei sedices e il mio sedicesimo anno
 TOR004: compleanno
 TOI054: compleanno
 TOI054: e mh
 TOI054: mh
 TOI054: sopra il mio al nostro alloggio dove avevamo la casa
 TOI054: e avev c’era una mansarda
 TOI054: to tutta unica e lui
 TOI054: e per un po’ di tempo ha diviso tutte ha fatte delle stanze poi aveva messo [80] la moquette avevamo messo [81] addirittura la tappezzeria
 TOI054: e io per il mio sedicesimo anno
 TOI054: avevo tutto e poi mi aveva comprato [82] lo stereo
 TOI054: e avevam fatto [83] la festa
- TOR004: did you have friends in the town you lived in
 TOI054: no me lots of friends
 TOI054: I always had lots lots of connections but lots of friends too lots of people that liked being with me
 TOR004: well but did you organize parties do things
 TOI054: because however
 TOI054: yeah also at home my dad for my 16th for my 16th year
 TOR004: birthday
 TOI054: birthday
 TOI054: and uhm
 TOI054: uhm
 TOI054: over mine our flat were we had the house
 TOI054: and we had there was an attic
 TOI054: all open and he
 TOI054: for a while he divided all he made rooms then he put (lit. *had put* [80]) the carpet we even put (lit. *had put* [81]) the wallpaper
 TOI054: and I for my 16th year
 TOI054: I had it all and then he bought (lit. *had bought* [82]) me the stereo
 TOI054: and we made (lit. *had made* [83]) the party

While the action of remembering was not mentioned in (14), it is clear that TOR004’s questions encouraged TOI054 to share memories. The adverbs *poi* (‘then’) could be interpreted as having had a listing function rather than a temporal one: indeed, while Pluperfects [80]–[82] necessarily preceded Pluperfect [83], the focus did not seem to be on the temporal collocation of each event with respect to the others, but rather on the totality of the elements that made up the memory of the party.

- (15) TOR004: e non sei mai andata all’universita’
 TOI054: no
 TOI054: si’
 TOI054: scherzando andavo

TOR004: ah
 TOR004: in che senso
 TOI054: del tipo
 TOI054: che andavo a scuola a ragioneria
 TOR004: eh
 TOI054: a cirie'
 TOI054: prendevo il treno con la mia amica
 TOR004: okay
 TOI054: e andavamo a torino
 TOI054: e poi andavamo all'universita'
 TOI054: ed e' successo d ascoltare anche delle lezioni
 TOR004: delle lezioni di cosa
 TOI054: e avevamo ascoltato [79] delle lezioni di biologia
 TOR004: ah
 TOI054: e poi prendevamo il quaderno con degli appunti facevamo
 TOI054: facevamo le le universitarie

TOR004: and you never went to university
 TOI054: no
 TOI054: yes
 TOI054: I went as a joke
 TOR004: oh
 TOR004: what do you mean
 TOI054: like
 TOI054: I went to high school
 TOR004: uhm
 TOI054: in Ciriè
 TOI054: I took the train with my friend
 TOR004: okay
 TOI054: and we went to Torino
 TOI054: and then we went to the university
 TOI054: and it happened that we listened lessons too
 TOR004: what lessons
 TOI054: well we listened (lit. *had listened* [79]) biology lessons
 TOR004: oh
 TOI054: and we took the notebook with notes we played
 TOI054: we played university students

The event encoded by Pluperfect [79] might be considered a memory too; nevertheless, the main reason why it belongs to a time frame in the past, and closed, is that, as it appears by reading a broader section of the conversation, it did not have any consequence on TOI054's life (e.g., she did not enroll in a biology course afterwards). Interestingly, had a Present Perfect been used instead, it would have received an experiential interpretation, i.e., it would have been considered to indicate that the situation, of having heard biology lectures, occurred at least once during a period of time extending to the present (Comrie 1976). While there is no doubt that the situation did indeed occur, the use of the Pluperfect seems to suggest (in contrast to the Present Perfect) that it was nonetheless of little significance to the speaker. This will be further discussed in Section 4. In more than one case, the events encoded by the past-temporal-frame Pluperfects displayed quite specific characteristics.

- (16) TOR002: pensa che dove c'e' adesso l'areoporto di caselle mio nonno aveva un terreno che gli hanno espropriato
 TOI119: si'
 TOI119: eh
 TOR002: quando han costruito l'areoporto nuovo

TOI118: mh
 TOI119: ah si' ma poi era caduto [17] anche l'aereo la' nelle case
 TOI118: caselle
 TOR002: eh si'

TOR002: think that where there now is the Caselle airport my grandpa had land that they expropriated
 TOI119: yes
 TOI119: uhm
 TOR002: when they built the new airport
 TOI118: uhm
 TOI119: oh well but then even the plan fell (lit. *had fallen* [17]) there in the houses
 TOI118: Caselle
 TOR002: yeah

In (16), the event encoded by Pluperfect [50] had two characteristics: on the one hand, as suggested by the preceding *ah si ma poi* 'oh well but then', it discursively appeared as a digression. On the other hand, given that TOI118 did not comment on the matter, and TOR002 simply answered with *eh si* 'yeah', it was also shared knowledge between the speakers.

- (17) TOI077: minchia l'ho portata in camper
 TOI077: gia' che
 TOI077: saliva in camper
 TOI077: poi siamo arrivati a sto posto li'
 TOR004: non era
 TOI076: eh eh la racconto io
 TOR004: non eri convinta
 TOI076: no
 TOI077: ma per niente aveva paura voleva andarsene via
 TOI076: ah gia' e' vero avevo chiamato [46] mia mamma
 TOI076: mentre tu eri sceso a parlare
- TOI077: shit I brought her camping
 TOI077: already
 TOI077: getting on the camper
 TOI077: then we arrived in that place there
 TOR004: it wasn't
 TOI076: uh uh I tell it
 TOR004: you weren't convinced
 TOI076: no
 TOI077: not at all she was scared she wanted to leave
 TOI076: oh right that's true I called (lit. *had called* [46]) my mom
 TOI076: while you had gotten off to talk

The event encoded by Pluperfect [46] seemed to come with a sense of surprise, on the part of the speaker, in recalling the event itself. In Section 4, it will be argued that these findings should be considered by further research addressing the hypothesis of the Pluperfect having additional functions to those analyzed in this paper.

Not all past-temporal-frame Pluperfects belong to a remote past, as one might assume, based on the previous examples (remember that the corpus was unbalanced in favor of remote contexts, as mentioned in Section 2):

- (18) TOI051: delle maschere ho fatto tanti di quei vestiti
 TOI052: tre
 TOR007: anche mia nonna uguale

TOI052: bellissimi
 TOI052: bellissimi proprio
 TOI052: davvero
 TOI052: anche
 TOI052: quello che *avevi fatto* [120] vedere l'altro giorno di quando mamma ha fatto la
 TOI051: ah la danzatrice

TOI051: some masks I made so many dresses
 TOI052: three
 TOR007: also my grandma the same
 TOI052: very beautiful
 TOI052: very beautiful really
 TOI052: for real
 TOI052: also
 TOI052: the one that you showed (lit. *had showed* [120]) the other day of when mom was
 TOI051: oh the dancer

Indeed, the temporal location of Pluperfect [120] was *l'altro giorno* 'the other day', i.e., quite recently in the past. This will be further discussed in Section 4.

3.5. Left-Out Occurrences

Six Pluperfects could not be assigned to either of the four functions: in two of these cases, the events encoded by the Pluperfects seemed to be digressions in discourse:

- (19) TOR004: e com'e' che siete finiti la'
 TOI054: non lo so
 TOI054: da questa superstrada che dava la cartina
 TOI054: oltretutto ero andata [87] con aldo eh
 TOR004: pensa te
 TOI054: e quindi lui era uno che
 TOI054: sapeva girare nel senso guardare la cartina non era un
 TOI054: uno che si perdeva ecco
- TOR004: and how is it that you ended up there
 TOI054: I don't know
 TOI054: from this highway the map said
 TOI054: besides I went (lit. *had gone* [87]) with Aldo uh
 TOR004: just think
 TOI054: and so he was one that
 TOI054: he knew how to travel I mean look at the map he wasn't a
 TOI054: one that got lost okay

Note that (19) is from the same conversation as (11). By reading the whole conversation, Pluperfect [87] appears to be deictic, and (19) seems to be a past-temporal-frame prototypical context (TOI054 was recalling the past experience of a road trip). Nevertheless, no other Pluperfects were used in recalling the trip. 'Having gone with Aldo' does not seem to be a more past-bound element than the others, but rather accessory information, added as it came to TOI054's mind—not different, in this aspect, from Pluperfect [50] in (16) (the latter, however, still displayed past temporal frame semantics, and had therefore been categorized as such, as explained in the previous section).

On the other hand, the other four left-out occurrences displayed a greater difference from the other Pluperfects analyzed so far (see above), in that they also seemed to be deictic,

but did not appear to be digressions, and not even the context they belonged to was past temporal frame.

- (20) TOI076: perche' abbi^{am} fatto ho fatto diverse lampade o anche lampadari
 TOI076: eh queste si' ci mi piacciono mi piacciono tanto
 TOR004: eh come fate per fare i lampada cioe' dovete farvi tutto lo studio dei
 TOR004: dei cavi
 TOI076: si'
 TOI076: so fare collegamenti elettrici io eh
 TOI077: eh quello che gli *avevo insegnato* [27] io
 TOI076: ho imparato da
 TOI076: da giulio anche eh dal mio suocero
 TOI077: si' anche mio padre

TOI076: because we made I made many lamps or also chandeliers
 TOI076: uh these yes we I like I like these a lot
 TOR004: uhm how do you do to make chandeliers I mean you have to study all
 TOR004: the cables
 TOI076: yes
 TOI076: I know how to make electrical connections duh
 TOI077: well that that I taught (lit. *had taught* [27]) her
 TOI076: I learnt from
 TOI076: from Giulio also uhm from my father-in-law
 TOI077: yes also from my father

It may have been the case that TOI077 was remembering the time he taught TOI076 how to make electrical connections, but this would not justify a past temporal reading of the event, given that its consequences were still extremely relevant (TOI076 knew how to make electrical connections). The relevance at the ST of the event encoded by the Pluperfect is perhaps more noticeable in (21), where TOI052 was showing TOI051 (her grandmother) videos that she had received:

- (21) TOI052: guarda nonna ti faccio vedere
 TOI052: eh
 TOI052: marco
 TOI051: si'
 TOI052: la bimba e' cresciuta guarda eh qui
 TOI051: uh uh
 TOI052: gli ha fatto una canzone sai che suona marco
 TOI051: certo lo so
 TOI051: guarda guarda x com'e' attenta
 TOI052: guarda qui bella
 TOI051: guarda co
 TOI051: ma che cara
 TOI052: si' son e' bellissima
 TOI052: eh
 TOI052: e invece giulia *aveva mandato* [125]
 [...] (addressees don't listen as they are still commenting on the video of the song)
 TOI052: e invece giulia
 TOI052: ha mandato
 TOI052: si sente il cuoricino di adele aspetta
- TOI052: look grandma I'll show you
 TOI052: uh
 TOI052: Marco

TOI051: yes
 TOI052: the baby has grown look here
 TOI051: uh uh
 TOI052: he wrote her a song you know that Marco plays
 TOI051: of course I know
 TOI051: look look how she's alert
 TOI052: look here pretty
 TOI051: look how
 TOI051: she's so sweet
 TOI052: yes they she's very beautiful
 TOI052: uhm
 TOI052: and Giulia send (lit. *had sent* [125]) instead
 [...] (addressees don't listen as they are still commenting on the video of the song)
 TOI052: and Giulia instead
 TOI052: has sent
 TOI052: you can hear Adele's little heart wait

Clearly, the fact that Giulia had sent TOI052 a video was still relevant, given that the video was about to be played. It is interesting to note that when TOI052 repeated the information, she switched to a Present Perfect (*ha mandato* 'has sent').

(22) TOI052: e tra l'altro il nonno quando
 TOI052: lui non c'e' quando si fanno i compleanni pero' se lo chiami la canzoncina
 te la canta sempre
 TOI051: no l
 TOI051: si' si' e' vero
 TOI051: l lo canta anche per telefono eh
 TOI052: certo si' infatti lo chiami
 TOI051: eh perche'
 TOI052: quest'anno tra l'altro vabbe' quest'anno ci siamo visti quindi alla fine non
 mi aveva chiamato [111]
 TOI051: si' eravamo li' eh

TOI052: and by the way grandpa
 TOI052: he doesn't come when we celebrate birthdays but if you call him he always
 sings you the song
 TOI051: no
 TOI051: yes yes it's true
 TOI051: he also sings it over the phone duh
 TOI052: of course yes you call him indeed
 TOI051: uhm because
 TOI052: this year by the way well this year we saw each other so in the end he
 didn't call (lit. *hadn't called* [111])
 TOI051: yes we were there uh

In (22), the temporal frame to which Pluperfect [111] belongs is undoubtedly still open, i.e., it includes the ST, as signaled by the time indication *quest'anno* 'this year'. The expression *tra l'altro* 'by the way' could indeed suggest that the event encoded by Pluperfect [111] was a digression; nevertheless, it could also be interpreted as shared knowledge between the speakers, given that TOI051 (TOI052's grandmother) confirmed that her husband (TOI052's grandfather) and she were with TOI052 on her last birthday.

(23) TOR002: io guido si chiama io guido
 TOR001: bravo
 TOI012: bravissimo
 TOI013: si'

TOI012: e' vero

TOR001: e' iniziato molto pr perche' io mi ricordo che quando *ero venuta* [227] qua a torino c'erano gia' e a milano no

TOR002: Io guido it's called Io guido

TOR001: bravo

TOI012: bravissimo

TOI013: yes

TOI012: it's true

TOR001: it began very early because I remember that when I arrived (lit. *had arrived* [227]) here in Torino there were already but not in Milano

In (23), car sharing services were being discussed. Despite being deictic, Pluperfect [227] encoded an event belonging to a temporal frame still open, given that the adverb *qua* 'here' suggests that TOR001 was still in Torino.

4. Discussion

The four temporal–aspectual functions described by Squartini (1999) (past-in-the-past, perfect-in-the-past, reversed result and past temporal frame) have proven to be indeed relevant for a description of spoken Italian, given that they managed to account for 239 of the 245 Pluperfect occurrences. The use of authentic language samples has allowed a further description of the aforementioned categories and of their prototypical context or context of use:

- the past-in-the-past function (145/245) is used to temporally organize events with respect to one other, i.e., to locate the event encoded by the Pluperfect prior to another past event (which can also be a proper consequence of the former). The temporal collocation of the events may be further specified by the presence of adverbial modifiers (e.g., *una volta prima* 'one time before').
- The perfect-in-the-past function (28/245) is used to highlight the relevance of the event encoded by the Pluperfect at a later past time. This reading is naturally compatible with the adverb *gia'* 'already', and with expressions measuring the temporal distance of the event encoded by the Pluperfect to the RT.
- The reversed result function (16/245) is used to stress that the results of the event encoded by the Pluperfect have been reversed at a later time in the past. This reading can often be confirmed by a following sentence describing the reversed situation that holds at the ST (eventually introduced by adverbs such as *invece* 'instead' or *poi* 'then').
- The past temporal frame function (63/245) is used to stress that the event encoded by the Pluperfect is past-bound. It is often used in contexts of remembering, and may co-occur with expressions with the noun *volta* 'time' (e.g., *la volta* 'the time (that)', *la prima volta* 'the first time (that)', etc.).

The reversed result and past temporal frame functions can be understood as instances of discontinuous past marking. Plungian and van der Auwera (2006) define the meaning of discontinuous past markers as “past with no present relevance” or “past and not present”, and explicitly state that the terms they use to distinguish these two subtypes (i.e., canceled result and framepast) are close to those employed by Squartini (1999) to refer to the Pluperfect’s two derived values. This suggests that the notion of discontinuity (Plungian and van der Auwera 2006) may be relevant for a description of the Italian verb system. Furthermore, it was shown in Section 3 that reversed-result Pluperfects (and past-temporal-frame Pluperfects, occasionally) can shape their meaning in opposition to the main kinds of Perfects (i.e., Perfect of result, Perfect of persistent situation and experiential Perfect). This suggests that discontinuity, while being considered purely temporal by Plungian and van der Auwera (2006), could be researched in the future as a double-faced notion. In fact, while the notion of past temporal frame can easily be interpreted as temporal, the reversed result notion seems to be closely related to aspect in its ‘anti-Perfect’ meaning.

Concerning the hypothesis of the existence of an aoristic use of the Pluperfect, the deictic Pluperfect occurrences that could not be classified as instances of the past temporal frame function (the only deictic function among those analyzed) were not only too small in number to prove the existence of an aoristic use, but also displayed different main connotations (i.e., discourse digression, shared knowledge) that also arose as secondary meanings amongst past-temporal-frame Pluperfects (in addition to surprise on behalf of the speaker). In fact, the typological literature has already highlighted that one of the most common derived uses of the Pluperfect is the marking of background information (e.g., digressions) (Plungian and van der Auwera 2006), and literature on Italian (Bertinetto 1986; Maiden and Robustelli 2007) has confirmed the existence of this use, albeit apparently considering it as arising from the original anaphoric meaning of the Pluperfect, which is considered to be preserved. On the other hand, according to Bermúdez (2011), the Castilian Pluperfect can be used as a marking of evidential distance, which includes shared access to the source of information (i.e., shared knowledge) and mirativity (i.e., surprise on behalf of the speaker). It is possible that a number of Pluperfects in the data encode new functions yet to be identified by exploring the domains of discourse and evidentiality, which have been subsumed under the past temporal frame category—probably also due to ParlaTO displaying a lot of past temporal frame contexts itself.

While further research is needed to identify the precise semantic scope of the Italian Pluperfect, its less prototypical functions (reversed result and past temporal frame), and the semantic values that arise in uncategorized data and/or in the past-temporal-frame Pluperfects, are all, in one way or another, based on the speaker’s perspective and/or on the speaker’s attention to the addressee. In principle, both reversed result and past temporal frame functions can be considered subjective in a proper semantic sense, as they both involve the speaker’s deictic center. In fact, although only the past temporal frame function is properly deictic, the reversed result function expresses a (reversed) resulting state that holds at ST—see (10), repeated in (24):

- (24) TOI003: ha perso tantissime cose torino
 TOR001: mh
 TOI003: se uno pensa
 TOI002: no pero’ si e’ arricchita parecchio con le olimpiadi a pa guarda prima non c’era
 TOI003: a partire da esperimenta
 TOI003: a partire da un macello di cose che io mi ricordo quando andavo a scuola potevi fare
 TOI003: un casi *era diventata* [187] la citta’ delle delle mh
 TOI002: mado’ ma prima tu vede mado’ ma tu prima vedevi turismo a torino
 TOI003: del libro e poi l’ha spostata a milano
- TOI003: Torino lost a lot of things
 TOR001: uhm
 TOI003: if one thinks
 TOI002: no but it developed a lot with the Olympics look before there wasn’t
 TOI003: starting from Esperimenta
 TOI003: starting from a lot of things that I remember when I still went to school you could do
 TOI003: a lot it became (lit. *had become* [187]) the city of of uhm
 TOI002: God but before you saw God but before you saw tourism in Torino
 TOI003: of books and then they moved it to Milano

In (24) the reversed resulting state of the event encoded by Pluperfect [187] was relevant to the speaker’s deictic center: Turin was not the city of books at the ST. In addition to being deictic, past-temporal-frame Pluperfects often occur in digressions, as in (16), repeated in (25):

- (25) TOR002: pensa che dove c'è adesso l'areoporto di caselle mio nonno aveva un terreno che gli hanno espropriato
 TOI119: si'
 TOI119: eh
 TOR002: quando han costruito l'areoporto nuovo
 TOI118: mh
 TOI119: ah si' ma poi era caduto [17] anche l'aereo la' nelle case
 TOI118: caselle
 TOR002: eh si'

TOR002: think that where there now is the Caselle airport my grandpa had land that they expropriated
 TOI119: yes
 TOI119: uhm
 TOR002: when they built the new airport
 TOI118: uhm
 TOI119: oh well but then even the plan fell (lit. *had fallen* [17]) there in the houses
 TOI118: Caselle
 TOR002: yeah

In (25), the speaker's perspective on the textual relevance of the event was at stake. Subjectivity was then involved, from at least two points of view: a more strictly semantic one and a textual one. The speaker's perspective was also in focus in cases in which his own surprise was highlighted, as in (17), repeated in (26):

- (26) TOI077: minchia l'ho portata in camper
 TOI077: gia' che
 TOI077: saliva in camper
 TOI077: poi siamo arrivati a sto posto li'
 TOR004: non era
 TOI076: eh eh la racconto io
 TOR004: non eri convinta
 TOI076: no
 TOI077: ma per niente aveva paura voleva andarsene via
 TOI076: ah gia' e' vero avevo chiamato [46] mia mamma
 TOI076: mentre tu eri sceso a parlare
- TOI077: shit I brought her camping
 TOI077: already
 TOI077: getting on the camper
 TOI077: then we arrived in that place there
 TOR004: it wasn't
 TOI076: uh uh I tell it
 TOR004: you weren't convinced
 TOI076: no
 TOI077: not at all she was scared she wanted to leave
 TOI076: oh right that's true I called (lit. *had called* [46]) my mom
 TOI076: while you had gotten off to talk

On the other hand, the shared knowledge value that arose in (25) can be understood as intersubjective, given that it required the speaker's attention to be conveyed towards the (alleged) knowledge of the addressee.

It appears that the notion of (inter)subjectification might be crucial in structuring a more integrated description of the less prototypical uses of the Italian Pluperfect, and in explaining its grammaticalization of new functions over time.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The KIParla corpus (of which ParlaTO is a module) is available at <https://kiparla.it/>, accessed on 12 April 2021.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- 1 Traugott (2003, 2010) uses the term (inter)subjectification to refer to the semanticization of subjectivity (the expression of the speaker's perspective and attitudes) and intersubjectivity (the expression of the speaker's attention to the self of the addressee).
- 2 This becomes particularly clear when one considers that a speaker of Northern Italian, referring to a deceased person, would not say *è nato* (lit. 'he has been born', i.e., using a present perfect which, in the variety under analysis, also encodes an aorist aspect) but *era nato* (lit. 'he had been born').
- 3 It is a module of the larger KIParla corpus (Mauri et al. 2019).
- 4 An effort has been made to provide translations as close as possible to the Italian texts, preserving the characteristics of the spoken language where possible.
- 5 The celebration of Turin's patron saint, St. John, that usually consists of a firework display on the river Po.
- 6 The sentences concerned would look as follows:
e non potevamo passare al secondo biennio se non avevamo già dato quattro esami obbligatori del primo
"and we couldn't move to the third and fourth years if we hadn't *already* passed four mandatory exams of the first two years"

io ne avevo già dati due, gli altri due erano enormi
"I had *already* passed two, the other two were huge"
- 7 'x's stand for incomprehensible text.
- 8 The transcription displays *vuoi* (want.PRS.2.SG), but from listening to the audio track it appears that TOI077 says *vuol* (want.PRS.3.SG) instead.
- 9 One can imagine that the complete sentence resembled either *avevo già capito che il passaggio era solo una scusa* ('I had *already* understood that the ride was just an excuse') or *avevo capito che le servisse davvero un passaggio* ('I had understood that she genuinely needed a ride').
- 10 The original (Perfect) meaning of the Italian Present Perfect is being (considered here, albeit it has come to encode an aoristic aspect too, especially in Northern Italy).
- 11 The sentence would look as follows:
Torino è diventata la città del libro
'Turin has become the city of books'.

References

- Bermúdez, Fernando Wachtmeister. 2011. El pluscuamperfecto como marcador evidencial en castellano. In *Estudios de tiempo y espacio en la gramática española*. Edited by Carsten Sinner, Elia Hernández Socas and Gerd Wotjak. Frankfurt am Main: Peter Lang, pp. 43–62.
- Bertinetto, Pier Marco. 1986. *Tempo, aspetto e azione nel verbo italiano: Il sistema dell'indicativo*. Firenze: Accademia della Crusca.
- Bertinetto, Pier Marco. 2003. *Tempi verbali e narrativa italiana dell'Ottocento: Quattro esercizi di stilistica della lingua*. Alessandria: Edizioni dell'Orso.
- Bertinetto, Pier Marco. 2014. Non conventional uses of the Pluperfect in Italian (and German) literary prose. In *Evolution in Romance Verbal Systems*. Edited by Emmanuelle Labeau and Jacques Bres. Bern: Peter Lang, pp. 145–70.
- Bertinetto, Pier Marco, and Mario Squartini. 1996. La distribuzione del Perfetto Semplice e del Perfetto Composto nelle diverse varietà di italiano. *Romance Philology* 49: 383–419.
- Cerruti, Massimo, and Silvia Ballarè. 2020. ParlaTO: Corpus del parlato di Torino. *Bollettino dell'Atlante Linguistico Italiano (BALI)* 44: 171–96.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- De Mauro, Tullio, Federico Mancini, Massimo Vedovelli, and Miriam Voghera. 1993. *Lessico di Frequenza dell'Italiano Parlato*. Milano: Etas.
- Diewald, Gabriele. 2011. Grammaticalization and pragmaticalization. In *The Oxford Handbook of Grammaticalization*. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press. [CrossRef]
- Hopper, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

- Maiden, Martin, and Cecilia Robustelli. 2007. *A Reference Grammar of Modern Italian*, 2nd ed. London: Hodder Arnold.
- Mauri, Caterina, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti, and Francesco Suriano. 2019. KIParla corpus: A new resource for spoken Italian. Paper present at the Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it, Bari, Italy, November 13–15. Edited by Raffaella Bernardi, Roberto Navigli and Giovanni Semeraro.
- Plungian, Vladimir, and Johan van der Auwera. 2006. Towards a typology of discontinuous past marking. *Language Typology and Universals* 59: 317–49. [CrossRef]
- Scarpel, Sebastiano. 2017. Considerazioni sull'uso aoristico del trapassato prossimo. *Studia de Cultura* 9: 121–31. [CrossRef]
- Squartini, Mario. 1999. On the semantics of the Pluperfect: Evidence from Germanic and Romance. *Linguistic Typology* 3: 51–89. [CrossRef]
- Traugott, Elizabeth Closs. 2003. From subjectification to intersubjectification. In *Motives for Language Change*. Edited by Raymond Hickey. Cambridge: Cambridge University Press, pp. 124–40.
- Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (Inter)subjectification: A Reassessment. In *Subjectification, Intersubjectification and Grammaticalization*. Edited by Kristin Davidse, Lieven Vandelanotte and Hubert Cuyckens. Berlin and New York: Mouton de Gruyter, pp. 29–71.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Diverging Grammaticalization Patterns across Spanish Varieties: The Case of *perdón* in Mexican and Peninsular Spanish

Marlies Jansegers^{1,*}, Chantal Melis² and Jennie Elenor Arrington Báez¹¹ Department of Linguistics, Ghent University, 9000 Ghent, Belgium; jennieleenor.arringtonbaez@ugent.be² Institute of Philological Research, The National Autonomous University of Mexico, Ciudad de México 04510, Mexico; cme@unam.mx

* Correspondence: marlies.jansegers@ugent.be

Abstract: This study investigates the contemporary grammaticalized uses of *perdón* ('sorry') in two varieties of Spanish, namely Mexican and Peninsular Spanish. Methodologically, the investigation is based on a taxonomy of offenses, organized around the concept of 'face' and based on spoken data of Spanish from Mexico and Spain. This taxonomy turns out to be a fruitful methodological tool for the analysis of apologetic markers: it does not only offer usage-based evidence for previous theorizing concerning the grammaticalization process of apologetic markers, but also leads to a refinement of these previous results from a contrastive point of view. Evidence from both corpora suggests a more advanced stage in the grammaticalization process of *perdón* in Mexican Spanish, where it can be used not only as a self-face-saving device geared towards the positive face of the speaker, but also in turn-taking contexts oriented towards the negative face of the interlocutor. Peninsular Spanish, on the other hand, resorts to a more varied gamut of apologetic markers in these contexts.

Keywords: grammaticalization; apologetic markers; politeness; *face*; offenses; Mexican and Peninsular Spanish; *perdón*

Citation: Jansegers, Marlies, Chantal Melis, and Jennie Elenor Arrington Báez. 2024. Diverging

Grammaticalization Patterns across Spanish Varieties: The Case of *perdón* in Mexican and Peninsular Spanish.

Languages 9: 13. <https://doi.org/10.3390/languages9010013>

Academic Editor: John Lipski

Received: 20 October 2023

Revised: 12 December 2023

Accepted: 15 December 2023

Published: 25 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The speech act of apology can rightly be considered “one of the most profound human interactions” (Lazare 2004, p. 1). This is reflected in the omnipresence of apologetic markers in our everyday life: consider for example the number of daily messages we start with “sorry to bother you” or “sorry for my late reply”, but also the vital importance of corporate apologies in customer service “sorry for the inconvenience” and of course, the recent avalanche of public apologies from governments, organizations and celebrities we see several times each week in the media. This pervasiveness of apologies has even led historians, philosophers, political and social scientists to coin the current era as an “Age of Apology” (cf., among others, Brooks (1999) and Gibney et al. (2007)).

From a linguistic point of view, these expressions such as (*I am*) *sorry* in English, *scusa* in Italian, *pardon* in French and *perdón* in Spanish have been qualified cross-linguistically in terms of grammaticalization, where the bleaching of the original semantic load gave rise to more procedural meanings (Molina 2011; Ghezzi and Molinelli 2019; Denoyelle 2020; Brenes Peña 2021). The contemporary Spanish marker *perdón*, for example, displays an ample gamut of uses that go beyond the original speech act of apology triggered by a previous offense and enter the realm of more procedural meanings related to various discourse-related functions. Consider the following examples:

- (1) I: empezaban en los Andes// en los Alpes *perdón*/ los Andes están acá abajo/ en los Alpes// y entonces/ venían desde Francia/ caminando/ y está el camino de Santiago. (CSCM)¹
 'I: they started in the Andes/// in the Alps sorry/ the Andes are down here/ in the Alps// and then/ they came from France/ walking/ and there is the road to Santiago'.
- (2) entonces yo digo <silencio/> *perdón* yo digo que este <vacilación/> cuando yo estaba por en <palabra_cortada/> entrar a la prepa era más fácil [...] (PRESEEA)
 'then I say <silence/> sorry I say that this <hesitation/> when I was about to <unfinished word/> enter high school it was easier'.
- (3) R: [(xx) eh: eh don Jesús *perdón* que lo interrumpa un segundito nada más tenemos a Abraham Mendoza en la línea estamos al aire para Panorama Informativo pues comentar específicamente lo que está sucediendo afuera y cómo se va a controlar (pues) esto parece ser incontrolable pero lo dejo al teléfono (CME)
 R: '[(xx) eh: eh don Jesús sorry to interrupt you for a second, we have Abraham Mendoza on the line we are on the air for Panorama Informativo to comment specifically on what is happening outside and how it is going to be controlled (well) this seems to be uncontrollable but I'll leave you on the phone'.
- (4) I: un rato llegué como a las/ dos y media de la mañana/ y me sacaron la bala hasta como por las nueve de la mañana
 E: ¿como a qué hora? *perdón*
 I: hasta las nueve (CSCM)
 'I: for a while I arrived at about 2:30 in the morning/ and they took the bullet out of me until about 9:00 in the morning.
 E: like at what time? sorry
 I: only around nine o'clock'
- (5) VV2F7 [v] Es que Johanson es la puta ama directamente VY2F8 [v] Esa mujer (0)
 VV2F7 [v]Luego está Chris Evans que para mí el capitán a -, a mí sí. VY2F8 [v] *Perdón* pero no. Para mí el el más guapo el mejor lo que sea es Thor. Lo siento mucho. VV2F7 [v] Bueno también Thor no está mal no te voy a engañar (CORMA)
 'VV2F7 [v] Is that Johanson is the fucking mistress directly VY2F8 [v] That woman (0) VV2F7 [v]Then there's Chris Evans who to me Captain a -, to me yes. VY2F8 [v] Sorry but no. For me the most handsome the best whatever is Thor. I'm sorry about that. VV2F7 [v] Well also Thor is not bad I'm not going to fool you'.

In example (1) *perdón* introduces a correction, whereas in (2) it is used to maintain the discursive thread. Moreover, *perdón* can also be used in contexts for turn-taking (3) or to request clarification (4) and even to attenuate upcoming criticism or difference of opinion (5).

Indeed, *perdón* has been related to multiple discourse-related functions. Fuentes Rodríguez (2009) includes *perdón* in her *Diccionario de conectores y operadores del español*, distinguishing between a “modal operator”, used to attenuate the illocutionary force of speech acts threatening to the interlocutor (a request, a refusal, an interruption, etc.), and a “connective reformulation marker”, appealed to in contexts of discourse self-repairs. Similarly, Brenes Peña (2021) organizes the procedural meanings of *perdón* developed by the original apologetic form along three dimensions: (1) metadiscursive, (2) interactional and (3) argumentative. She concludes that this leap from the sentence to the text as a unit of analysis of the lexeme *perdón* is a clear case of grammaticalization (or pragmaticalization, see below Section 2.2) where the bleaching of the original semantic load gave rise to more procedural meanings. From this perspective, then, *perdón* starts to compete with other markers displaying equivalent discourse-related functions such as *bueno* (disagreement, dispreferred responses and corrections), *oye/oiga* (to attract attention or act as a mitigator in controversial contexts), *o sea* (repair, utterance completion and clarification), *ahora* (disagreement marker), *este* (discourse flow, reformulation, hesitation), amongst many others.

However, if we identify *perdón* as a member of a paradigm that comprises other discourse markers (DM), the question arises as to what makes this marker unique compared with other (apparently) equivalent functional forms. The present study aspires to tackle this question and to discover the unique character of *perdón* by delving deeper into the grammaticalization process suffered by this apologetic marker. More precisely, some crucial questions remain unanswered:

- First, given the feature of persistence inherent to each grammaticalization process, what are the vestiges of the canonical model of apologies in the grammaticalized use of *perdón*?
- Second, if *perdón* is indeed the result of a grammaticalization process, to what extent can we relate its contemporary grammaticalized values to the original illocutionary meaning of the act of apology?
- Third, what are the underlying mechanisms and the subsequent compensatory forces for the bleaching of the original propositional illocutionary force of the original speech act of apology?

Taking into account the original semantics of *perdón* as well as the theoretical notions of *face* and *offense* crucial for the understanding of apologies, we propose an empirical study into the grammaticalization of contemporary *perdón* that aspires to both complement and deepen previous findings on this apologetic marker. Additionally, based on the idea that synchronic variation gives insight into ongoing change (Haverkate 1994; Wichmann et al. 2010; Winter-Froemel 2014; Lehmann [1995] 2015; Detges and Waltereit 2016; Gancedo Ruiz 2019) and the well-known fact that the use of politeness markers not only differs between languages, but also between varieties of the same language, we postulate the hypothesis that a comparison between two Spanish varieties might reveal diverging patterns of ongoing grammaticalization.

The outline of the paper is as follows: Section 2 provides a detailed theoretical description of the canonical model of apology (Section 2.1) and an overview of the previous studies on the origin and development of apologetic markers from a cross-linguistic point of view and the underlying mechanisms responsible for this semantic-pragmatic change (Section 2.2). Section 3 first discusses the method and data used in this study and then presents the results of a quantitative and qualitative analysis of the apologetic marker *perdón* in Mexican and Peninsular Spanish. Finally, in Section 4 these findings will lead to a detailed discussion on the contemporary uses of the marker, revealing diverging patterns of grammaticalization in both varieties.

2. Antecedents

2.1. The Canonical Model of Apologies

What all these contemporary apologetic formulas such as (*I am*) *sorry*, *scusa*, *pardon* and *perdón* have in common, is precisely their origin: they all developed from a speech act of apology. Therefore, in order to fully understand their contemporary behavior, it is crucial to first study the nature and essence of the canonical model of apologies.

Since Goffman (1971, pp. 138–48), the apology has been characterized as an essential element in a so-called *remedial interchange*. That is, the main purpose of an apology is to repair an offense committed by the speaker (the *offender*) against the interlocutor (the *offendee*) with the aim of restoring the social harmony or equilibrium between speakers (Edmondson 1981, p. 280; Leech 1983, p. 125). In the same vein, Holmes (1990, p. 159) provides the following definition of *apology*:

An apology is a speech act addressed to B's face-needs and intended to remedy an offense for which A takes responsibility, and thus to restore equilibrium between A and B (where A is the apologizer, and B is the person offended).

This definition clearly highlights three important theoretical concepts linked to the study of apologies, namely (1) the speech act of apology itself, (2) the concept of *face* and—

more generally—linguistic politeness theory and (3) the apology as a reactive speech act that implies a previous offense for which the speaker takes (at least partial) responsibility.

First of all, according to the canonical model of apologies, an apology can be realized either explicitly or implicitly (cf., among others, Olshtain and Cohen 1983; Blum-Kulka et al. 1989; Holmes 1990; Trosborg 1995; Márquez-Reiter 2000; Harris et al. 2006; González-Cruz 2012). Instances of the latter, indirect, strategies are for example giving an explanation (*Parece que hoy no me concentro en nada* ‘I just can’t seem to concentrate today’), an offer of redress (*Le puedo comprar uno nuevo* ‘I can buy you a new one’) or a promise of forbearance (*No volverá a ocurrir* ‘It won’t happen again’).² Alternatively, speakers can recur to explicit apologies which appear in the form of so-called illocutionary force indicating devices (IFIDs) as defined by Searle (1969, p. 62). For English, for example, the well-known taxonomy of apology IFIDs provided in the CCSARP coding manual (Cross Cultural Speech Act Realisation Patterns, cf. Blum-Kulka et al. 1989, p. 290) includes expressions with words such as *sorry*, *excuse*, *apologise* and *pardon*. Similarly, in Spanish, the IFIDs of apology can include both performative verbs such as *perdonar*, *disculpar*, *excusar*, *sentir*, *lamentar* (example 6) and formulas such as *lo siento (mucho) y perdón* (example 7):

- (6) Oh, *discúlpame* amor, se me pasó el tiempo. (CREA, 1983)
 ‘Oh, forgive me, honey, I lost track of time’.
- (7) Oh, *perdón* -dijo el comisario-. Ignoraba que fuese usted viudo. (CREA, 1975)
 ‘Oh, sorry, said the commissioner. I didn’t know you were a widower’.

Second, since Brown and Levinson’s (1987) seminal theory of politeness, apologies have been widely studied within the framework of politeness theory, reflecting as such the idea formulated by Holmes (1998, p. 217) that “the apology is quintessentially a politeness strategy”. Crucial in their theory is the concept of *face*, first adopted by Goffman (1967, 1971). Building on Goffman’s theory of face needs, Brown and Levinson (1987, p. 62) distinguish between *negative face* and *positive face*: the latter is defined as “the want of every member that his wants be desirable to at least some others”, that is, the human need for recognition and affection (belonging). Negative face, on the other hand, is defined as “the want of every ‘competent adult member’ that his actions be unimpeded by others”. This essentially refers to the human need to be autonomous. Other authors have (re)conceptualized negative face in Goffman’s (1967, p. 5) terms of *territory* and *territories of the self*, alluding either to bodily, material, spatial, temporal or cognitive territory (e.g., Kerbrat-Orecchioni 1992; Bravo 1999; Bello 2015).

Ideally, then, every human being would be at the same time unanimously appreciated (positive face) but left alone (negative face) by others. However, this is an unattainable ideal as almost any interaction involves acts that are potentially threatening to one, or both, of these types of *face*. Such acts that infringe on someone’s *face* have been labelled *Face Threatening Acts* (FTAs) by Brown and Levinson. In this view, then, the speech act of apology is considered to be face-saving for the hearer and face-threatening for the speaker (Olshtain 1989, p. 156). Indeed, as discussed above, an apology is typically aimed at face redress after committing an offense that has damaged the addressee’s face. This is illustrated, for example in sentences such as (6) above and (8), where the speaker apologizes for not respecting and interfering with the temporal territory of the hearer:

- (8) H2: Oye que *perdona* que hemos llegado tarde, ¿eh? pero es que... H4: [solapamiento de turnos] Ha sido culpa mía. (CORLEC)
 ‘Hey, forgive us for being late, huh? but... H4: [overlapping of turns] It was my fault’.

However, as Deutschmann (2003, p. 43) points out, considering negative face redress as the only function of apologies would be a far too narrow view of this “versatile speech act”. Indeed, as Holmes (1990, p. 162) already mentions, some apologies are geared towards positive face redress of the hearer. In the following example, the apology is supposed to redress an FTA to the hearer’s positive face (in this case forgetting someone’s name), explicitly attending to the hearer’s wants and needs to be recognized:

- (9) E: muy bien Raquel / ¿tú sabes cocinar? - I: Rosa Rosa <risas = "I"/ - E: ¡ay! *perdón* Rosa / ¿tú sabes cocinar? (PRESEEA)
 'E: very well Raquel / do you know how to cook? - I: Rosa Rosa <laughs = "I"/ - E: oh! sorry Rosa / do you know how to cook?'

Although most apology studies have focused on redress of hearer's (positive or negative) face, it is well known that human beings not only try to protect the face of others, but also aspire to safeguard their own face (Goffman 1972). In this respect, Chen (2001) coined the term *self-politeness*, whereas other scholars prefer to talk about activities related to *self-image* preservation, referring to those situations where a speaker fears that his/her image might be endangered in the eyes of others by some utterance or potentially harmful action, i.e., an FTA to the positive face of the speaker (cf., among others, Bravo 2005; Hernández Flores 2005, 2008). Examples include apologies for social gaffes such as coughing and slips of the tongue.

In line with Deutschmann (2003, p. 43), the function of apologies as politeness markers should thus be considered as multi-faceted in the literal sense: although traditionally their use has been regarded as a way to redress (positive or negative) hearer's face, in many cases the use of apology markers is also aimed at maintaining or improving the speaker's face, or self-image. The theoretical concept of *face* is thus a very useful tool to account for the complex nature of apologizing, as long as multiple *faces* possibly involved in the interaction are taken into consideration. Therefore, in the present study, these four different axes will be considered: both the negative and positive face needs of both the hearer and the speaker.

A third key element in the study of apology markers is their reactive nature, and more precisely, the presence of an offense for which the speaker takes (at least partial) responsibility. Indeed, the offense or "object of regret" (Coulmas 1981, p. 75) is what essentially motivates an apology. As discussed above, within the theory of Brown and Levinson, offenses are seen as FTAs that can be geared towards both the negative or positive face of the speaker or hearer. Interestingly, although there seems to be little consensus among scholars on an operational taxonomy of offenses resulting in apologies, they do coincide in that the nature and severity of the offense will, to a great extent, determine the form of the subsequent apology (cf., among others, Holmes 1990; Kerbrat-Orecchioni 1992; Aijmer 1996; Wagner 1999; Deutschmann 2003). For example, bumping into someone accidentally will result in a different apology than would breaking someone's new phone. That is, there is reason to suggest that the offenses motivating an apology can be ordered along a continuum that proceeds from more serious to minor offenses.

However, when apologetic markers are studied in context, the empirical reality turns out to be even more complex than this. Indeed, several authors (Goffman 1971; Norrick 1978; Coulmas 1981; Ogiermann 2009; Ghezzi and Molinelli 2019) observe that the use of apologetic markers does not always constitute "genuine" speech acts of apology. Not uncommonly, explicit forms of apologies are uttered when the offense is minimal or even non-existent. In these cases, apologizing is rather a matter of routine, a mere formula used for the purpose of complying with social norms of good or polite behavior. Interestingly, in these formulaic apologies, speakers are more likely to resort to short, ritualized formulas such as *sorry* in English or *perdón* in Spanish. Compared with the more elaborated performative verbs such as *perdonar* ('to forgive') and *disculpar* ('to apologize'), these formulas are partially stripped of their original illocutionary force of apology and thus semantically bleached. This is why these expressions have also been characterized in terms of *grammaticalization*, which will be explained in the next section.

2.2. Grammaticalization of Apologies

The contemporary apologetic marker *perdón* seems to have originated as an ellipsis from the expression *te pido perdón* (Fuentes Rodríguez 2009, p. 251). This origin is what Spanish *perdón* has in common with its counterparts in other languages such as Italian *scusa* (*ti chiedo scusa* > *scusa*), French *pardon* (*Je vous demande pardon* > *pardon*) and English

sorry (*I am sorry* > *sorry*) (Molina 2011; Ghezzi and Molinelli 2019). Besides their formal resemblance, these expressions also seem to share their contextual origin. From the very beginning, apologies are linked to a religious context and from there they have spread to civil society through a complex process of secularization. In line with Kohnen (2017) and Williams (2018), Jucker (2019a) traces the long diachrony of apologies in the history of the English language. He convincingly demonstrates how the Christian act of penitence and repentance should be considered as a precursor of the modern apology, or at least as the first step in the semantic development of apologies. The same impact of religion has also been identified for French in the study of Denoyelle (2020). Medina López (2023) has very recently shown a similar diachronic development for Spanish apologies: expressions such as *perdón*, *me arrepiento*, *mi culpa*, *lo lamento* y *lo siento* originally are speech acts with considerable weight addressed to God that through a process of attenuation are gradually desacralized and eventually become used as speech acts that are often no more than “a token acknowledgement of some minor infraction” (Jucker 2019a, p. 17).

This formal and semantic development of apologies has been studied from a wide variety of theoretical perspectives on diachronic change and an ample gamut of specific diachronic processes have been identified to describe the rise of apologetic markers. Norrick (1979) for example, considers English *sorry* and *pardon* to be lexicalized pragmatic formulas. Similarly, for Aijmer (1996), (*I am*) *sorry* is a case of lexicalization, although recognizing “degrees of lexicalization on a scale of frozenness” (Aijmer 1996, p. 10). In view of Molina (2011), *sorry* is a case of pragmaticalization, linked to grammaticalization, and for Jucker (2019a), this diachronic process can be best characterized in terms of speech act attenuation. Within the Spanish realm, Brenes Peña (2021, p. 142) analyzes the uses of *perdón* deliberately as a case of “gramaticalización o pragmatización” without choosing between either term. Indeed, this is of course just a terminological question to describe the same underlying semantic-pragmatic process. As Brinton (2017, p. 34) points out, the choice between terms like *grammaticalization* and *pragmaticalization* seems to hinge not so much on the process itself but rather on what is encompassed by “grammar”. The traditional conceptualization sees grammar as restricted to the morphosyntactic domain. Consequently, the process of grammaticalization focuses on the reduction of structure and form of linguistic units, while emphasizing the increase of their morphosyntactic dependence. If grammar is viewed from this rather narrow perspective, then pragmatic elements are excluded from the process of grammaticalization. Indeed, pragmatic markers may adopt textual functions, take care of discourse organization, express speaker stance or interpersonal values and can be used as politeness strategies, functions that “are not usually considered as the core business of grammar, if they are felt to be grammatical at all” (Van Bogaert 2011, pp. 315–16). However, different scholars have been arguing in favor of a more comprehensive, inclusive definition of *grammar* in order to go beyond the morphosyntactic level and also embrace discourse functions. As Traugott argues:

I see grammar as structuring communicative as well as cognitive aspects of language. Grammar encompasses phonology, morphosyntax, and truth-functional semantics, and is rich enough to license interaction with the general cognitive abilities such as are involved in the speaker-addressee negotiation that gives rise to grammaticalization. These include information processing, discourse management, and other abilities central to the linguistic pragmatics of focusing, topicalization, deixis, and discourse coherence. (Traugott 2003, p. 626)

Such a broad conceptualization of grammar allows for pragmatic markers to be incorporated into the realm of ‘grammar’ and, thus, to be studied from the perspective of grammaticalization as well. Of course, such a comprehensive conception of grammar makes a term like pragmaticalization dispensable. As a consequence, the diachronic development of several discourse and pragmatic markers have been described in terms of grammaticalization (Castillo Lluç 2008; Wichmann et al. 2010; Molina 2011; Hancil 2018).

Indeed, at first sight, some of the traditional parameters linked to grammaticalization are reflected in the semantic-pragmatic evolution of the Spanish apology marker *perdón* (cf. Hopper 1991):

- (i) *Layering*: because the forms *perdón*, *perdona/e*, *lo siento*, *lo lamento* coexist in the same functional domain.
- (ii) *Divergence*: because the full expression *te pido perdón por*, the lexical noun *perdón* and the apologetic marker *perdón* exist side by side.
- (iii) *Specialization*: according to previous studies (Fuentes Rodríguez 2009; Brenes Peña 2021) contemporary *perdón* specializes as a connective operating in three functional domains: (1) metadiscursive (2) interactional and (3) argumentative (see Introduction).
- (iv) *Persistence*: vestiges remain of the original (religious) meaning of *perdón* in the apologetic marker.
- (v) *Decategorialization*: loss of verbal/propositional characteristics.

However, the application of the parameters mentioned by Lehmann ([1995] 2015) is less straightforward: there is no obvious reduction of scope (condensation), no coalescence or morphological bonding, *perdón* does not become obligatory (obligatorification) nor paradigmaticized (paradigmaticization) in so far as it does not join a grammatical paradigm. Instead of scope reduction (condensation), *perdón*—just like other pragmatic markers—rather exhibits scope extension and positional freedom.

Importantly, strictly speaking, there is even no semantic bleaching involved. As we already know from the diachronic studies mentioned above, the origin of contemporary *perdón* is not a word or a phrase with lexical or propositional content, but a speech act in itself with illocutionary force: *perdón* originates as a speech act in penitential acts and confessions to God and later becomes gradually desacralized to expressions of forgiveness and subsequently regret. Nowadays, it can even be used as a mere formulaic speech act, often without the presence of a clear offense. The ultimate goal has always been to maintain harmonious relations with God or the interlocutor. That is, in its origin, we have a speech act that is basically oriented towards the hearer, essentially interpersonally.

In other words, apologetic markers such as *perdón* in Spanish and *sorry* in English are not instances of a change from propositional content towards speech acts, but rather embody a process that affects the speech act itself, and more precisely, its illocutionary potential. They have an illocutionary function already at the onset and this illocutionary material is recruited for further illocutionary and discursive uses. In order to recognize this phenomenon in its own right, Arnovick (1999) and Claridge and Arnovick (2010, p. 187) coined this process *discursisation*. More recently, Jucker (2019a) uses the term *attenuation* for this process, i.e., a progressive weakening of its illocutionary force, and proposes a scale of speech act attenuation that is not unique to apologies but has wider applications. Applied to the history of apologies, he defines this as follows:

The [...] history of apologies [...] shows a process of attenuation, that is to say their force becomes increasingly weaker, and at the same time the linguistic resources used to perform the speech act undergo a process of reduction and conventionalization. In its early form the speech act is semantically explicit and spells out its illocution while later it is reduced to a conventionalized expression that requires increased pragmatic processing for its interpretation. (Jucker 2019a, p. 7)

This “increased pragmatic processing” is what Claridge and Arnovick (2010) call “pragmatic strengthening”.

Based on this theoretical background, in the remainder of this paper, we will present the results of an empirical, comparative study of the uses of *perdón* in contemporary Spanish. By comparing two different varieties of Spanish (Mexican and Peninsular Spanish) we aspire to add new insights into this ongoing debate on the grammaticalization, discursisation or attenuation of apology markers.

3. Case Study: *Perdón* in Mexican and Peninsular Spanish

3.1. Methodology and Corpus

The data used to investigate the contemporary uses of the apology marker *perdón* come from a sizeable, manually annotated corpus consisting of spoken spontaneous conversations and interviews, recorded during the last quarter of the 20th century and the first decades of the 21st century. The following list of existing, spoken corpora were exploited: Corpus del Proyecto para el Estudio Sociolingüístico del Español de España y de América (PRESEEA), América y España español coloquial (AMERESCO), Corpus Sociolingüístico de la Ciudad de México (CSCM), Corpus del Habla de Baja California (CHBC) and Corpus Michoacano del Español (CME) for Mexican Spanish and Corpus Oral de Madrid (CORMA), Valencia Español Coloquial (Val.Es.Co.), Corpus Oral de Referencia del Español Contemporáneo (CORLEC), Corpus integrado de referencia en lenguas romances (C-ORAL-ROM) and Corpus del Proyecto para el Estudio Sociolingüístico del Español de España y de América (PRESEEA) for Peninsular Spanish. In a first phase we collected all occurrences of *perdón* together with its near-synonymous apologetic markers such as *lo siento* and forms derived from the performative verbs *perdonar* and *disculpar*. This first step allowed us to quantitatively assess the overall frequency and productivity of *perdón* compared with its near-synonymous expressions. Following this objective, we retrieved a total of 769 occurrences: 363 cases for Mexican Spanish and 406 for Peninsular Spanish. From this corpus, we then selected only the instances of *perdón*, yielding a total of 500 instances (299 for Mexico and 201 for Spain) that were subjected to a fine-grained qualitative analysis.

For this qualitative study, we analyzed the speech act of apology from the perspective of the offenses that motivate the apology and built on the hypothesis presented in the theoretical introduction (see Section 2.1) that the nature and severity of the offense partly determine the forms chosen to apologize. The analytical tool used is a taxonomy of offenses, organized around the concept of face as conceptualized by Goffman (1967) and Brown and Levinson (1987). As Deutschmann (2003, p. 62) already pointed out, unfortunately there has been little consensus over the taxonomy of offenses in past studies. Consequently, we decided to propose our own taxonomy of offenses in Spanish. This taxonomy is based on the empirical data from our own corpus and supported by a meticulous review of the literature on the subject. Table 1 below summarizes our taxonomy of offenses motivating apologies in Spanish, organized around the concept of face for both the speaker and the interlocutor:

Table 1. Taxonomy of offenses in Spanish, organized in terms of face.

| Negative Face (Hearer) | Positive Face (Hearer) | Positive Face (Speaker) |
|---|------------------------------------|---|
| Invasion of spatial territory | Criticism or disagreement | Slips of the tongue (<i>lapsus linguae</i>) |
| Interference with temporal territory | Lack of consideration | Censored language |
| Interruptions | Rejection of offers or invitations | Inappropriate behavior |
| Damage to belongings | Breach of promise | Social gaffes |
| Violation of the right not to be distracted | | |
| Obligation to do something | | |
| Obstruction of plans | | |

As Table 1 shows, this taxonomy embodies an attempt to relate, in a precise manner, each type of offense to the typical traits associated with the positive and negative face of both interlocutors. As such, it encompasses three broad classes, made up of several subclasses: (1) offenses that damage the hearer’s negative face (2) offenses aimed at the hearer’s positive face and (3) offenses that threaten the speaker’s positive face.³ Applied to our corpus data, this means that all 769 occurrences were analyzed manually, identifying for each of them, (a) the specific formula of apology (*perdón*, *perdonaje*, *disculpa/e*, *lo siento*), (b) the type of offense, (c) the face affected (positive face/negative face), and (d) the orientation of the face (speaker/hearer). In the following sections, we will first present

the overall quantitative results (Section 3.2) and then turn to the more in-depth qualitative analysis of the marker *perdón* (Sections 3.3 and 4).

3.2. Quantitative Results

First of all, the question arises as to what are the most commonly used explicit apology formulas in Spanish. Table 2 displays the overall frequency of each form in the Mexican and Peninsular corpus.

Table 2. Apology forms in Mexican and Peninsular Spanish. ($\chi^2 = 192$; $df = 3$, $p < 0.001$).

| Forms | Mexico | | Spain | |
|------------------|--------|------|-------|------|
| | # | % | # | % |
| <i>disculpar</i> | 49 | 14% | 12 | 3% |
| <i>perdonar</i> | 11 | 3% | 158 | 39% |
| <i>lo siento</i> | 4 | 1% | 35 | 9% |
| <i>perdón</i> | 299 | 82% | 201 | 49% |
| Total | 363 | 100% | 406 | 100% |

It is striking that, although all apology forms are present in both varieties, their distribution diverges considerably between Mexican and Peninsular Spanish. What both varieties have in common—and what immediately catches the eye, of course—is the clear supremacy of *perdón* in this pragmatic domain of Spanish apologetic markers. However, this dominance is much more evident in Mexican Spanish, where *perdón* covers 82% of all cases and outperforms by far all other forms in the apologetic domain. In Peninsular Spanish, on the other hand, there still seems to be a stronger competition between *perdón* (49%) and the performative verb *perdonar* (39%). These quantitative differences between both varieties possibly point towards some important underlying qualitative divergences related to the functional scope of the marker *perdón* in Mexican and Peninsular Spanish. Indeed, in light of Zipf’s (1949) *Principle of Economic Versatility* according to which frequency of use is correlated with semantic versatility, the higher frequency of *perdón* within the functional domain of apologetic markers might correlate with some semantic-pragmatic enrichment, eventually leading to a grammaticalized discursive formula. Additionally, the frequency differences between both varieties also suggest some dialectal variation in this grammaticalization process. In order to interpret these quantitative data, in the next section we turn to an in-depth contrastive analysis of the most frequent apology marker in both varieties, *perdón*.

3.3. *Perdón* in Mexican vs. Peninsular Spanish

In order to define the underlying mechanisms responsible for the frequency differences in both varieties, we analyzed each occurrence of the corpus for three qualitative variables, namely (a) the specific type of offense, (b) the face affected (positive/negative) and (c) the orientation of the face (speaker/hearer). Considering these variables, the distribution for the specific forms in the Mexican corpus is represented in Table 3:

Table 3. Relation between (orientation of) face and form of apology in Mexican Spanish.

| | <i>Disculpar</i> | | <i>Perdonar</i> | | <i>Lo Siento</i> | | <i>Perdón</i> | | Total | |
|-----------------|------------------|-----|-----------------|----|------------------|-----|---------------|-----|-------|------|
| | # | % | # | % | # | % | # | % | # | % |
| Positive face S | 15 | 7% | 3 | 1 | - | - | 196 | 92% | 214 | 100% |
| Negative face H | 22 | 18% | 6 | 5% | 1 | 1% | 90 | 76% | 119 | 100% |
| Positive face H | 12 | 40% | 2 | 7% | 3 | 10% | 13 | 43% | 30 | 100% |

The frequencies in this table reveal a very interesting profile of the marker *perdón* in the Mexican corpus. Compared with the other possible apologetic forms at the speaker’s

disposal, *perdón* turns out to be the default form for the protection of the speaker's positive face (92%), highly preferred after offenses geared towards the hearer's negative face (76%) and much less used for offenses towards the hearer's positive face (43%). As we saw in the taxonomy of offenses presented in Table 1, the first category of offenses towards the speaker's positive face encompasses slips of the tongue (example 10), censored language (example 11), inappropriate behavior (for example leaving a place abruptly, see example 12) or social gaffes such as sneezing or burping (example 13):

- (10) [...] o sea/ prácticamente toda la secundaria/ toda la prepa *perdón*/ vivía en su casa
 [...] (CHBC)
 'I mean/ practically all of secondary school/ all of high school excuse me/ I lived in his house [...]'
- (11) B: [sí otra] vez no/ no mamá cuando una gente es *perdón* pero <énfasis
 t="pronunciación_marcada">imbécil</énfasis>/ es imbécil toda su vida] (AMERESCO)
 'B: [yes again] no/ no mom when someone is sorry but <emphasis
 t="pronunciación_marcada">imbecile</emphasis>/ is imbecil all his life!'
- (12) A: y yo estoy bien así o sea <alargamiento/>/// con el respeto se basa porque <fsr
 t="pus">pues</fsr>/// luego tardo ocho días en verla// diez días quince días
 dependiendo/// (1.6) <fsr t="horita">ahorita</fsr> la vi ¿cuándo? ¿martes?/// (2.2) tal
 vez no la hubiera visto si hubiera trabajado no la hubiera visto/ hasta// el miércoles o el
 jueves/// (1.8) es así/// (1.2) nos mantenemos así respetados/ pero/// (1.8) <fsr
 t="horita">ahorita</fsr>se va a enojar si no voy <risas/>/// porque// es
 que<alargamiento/>/// (1) me retiro señora/ *perdón*
 C: ¿[sabe que estás] conmigo?
 B: [<ininteligible/>]
 A: le dije [que fui a mi]sa (AMERESCO)
 'A: and I'm fine like this, that is to say</>/// with respect it is based because <fsr
 t="pus">then</fsr>/// it takes me eight days to see her/// ten days fifteen days
 depending/// (1.6) <fsr t="horita">ahorita</fsr> I saw her when?/// Tuesday?///
 (2.2) maybe I wouldn't have seen her if I had worked I wouldn't have seen her/// until
 Wednesday or Thursday/// (1.8) it's like this/// (1.2) we keep each other respected that
 way/// but/// (1.8) <fsr t="horita">ahorita</fsr> she'll get mad if I don't go
 <laughs/>/// because/// it's that<lengthening/>/// (1) I'm leaving madam/ sorry.
 C: [does he know that you are] with me?
 B: [<unintelligible/>]
 A: I told him that I went to Mass'
- (13) I: fue muy / a mí me gustó ser *cúrpite* esa vez / y para las danzas me gustó mucho ser
 maringuilla // y fijate que hasta / para cuestión // <ruido = "eructo"/> *perdón* / de
 vestuario // (PRESEEA)
 'it was very / I liked being a *cúrpite* that time / and for the dances I really liked being a
maringuilla // and notice that even / for the question // <noise = "burp"/> sorry /
 of costumes'

As can be seen from the examples above, this category encompasses situations where the offense for the addressee is minimal, even non-existent. In these cases of "formulaic" apologies (Deutschmann 2003, p. 46), the presence of an IFID of apology seems to be rather a matter of routine, primarily motivated by the desire to "evince good manners" (Norrick 1978).

Besides the offenses towards speakers' positive face, *perdón* is also very frequent in contexts with offenses towards the hearer's negative face. In this category, *perdón* turns out to be extremely productive in the case of interruptions and turn-taking contexts (62 out of 90 occurrences, see example 14):

- (14) I: un rato llegué como a las/ dos y media de la mañana/ y me sacaron la bala hasta como por las nueve de la mañana
 E: ¿como a qué hora? *Perdón*
 I: hasta las nueve (CSCM)
 T: for a while I arrived at about 2:30 in the morning/ and they took the bullet out of me until about 9:00 in the morning.
 I: about what time? Sorry
 I: until nine o'clock'

These examples of turn-taking contexts illustrate that *perdón* is indeed moving towards discourse organization, which coincides with the interactional dimension distinguished by Brenes Peña (2021) in the grammaticalization process suffered by *perdón*.

Finally, compared with these two categories, the offenses oriented towards the hearer's positive face seem to somewhat resist the use of *perdón*. In these cases, *perdón* starts to compete with the more elaborate, performative verb *disculpar* (resp. 43% vs. 40%). This is the case, for example, with disagreements and criticisms, as illustrated in example (15):

- (15) eh nosotros ya ya veíamos que que la gente pues se limitaba exclusivamente a eso el turismo en sí no: no prolifera:ba no era pero le le vuelvo a insistir le vuelvo a insistir esto no *perdón* no lo veo yo como: que sea algo en su totalidad de de echado a perde:r de: de que quisiéramos que ya no vinieran no no no este esto es lo que no- nos da vida porque mucha gente se da cuenta de lo que realmente es- [CME]
 'eh we already saw that people were limited exclusively to that, tourism itself was not proliferating but I insist again I insist again this is not sorry I do not see it as something totally spoiled that we would not want them to come anymore no no no no hum this is what does not-it gives us life because many people realize what it really is'

Interestingly, as discussed above (see Section 2.1), some previous studies have suggested that the offenses motivating an apology can be ordered along a continuum that proceeds from more serious to minor offenses. Applying this rationale to our tripartite taxonomy of offenses, we can indeed arrange the three categories described above along this continuum. At the pole of serious offenses are threats to the positive face of the hearer, such as criticism or disagreement. These instances can be classified as more serious offenses on the grounds that they challenge the other's social dignity (his/her positive face-wants of belonging). Next in line are the offenses threatening the hearer's negative face, as they involve violations of one's claims to privacy and freedom from impositions and impediments. Compared with the previous category of serious offenses, these are minor offenses because in these circumstances the expression of apology does not express true regret on the part of the speaker but serves instead as a "token acknowledgment of some minor infraction" (Jucker 2019a, p. 17), "associated with the implied message 'please, don't think I'm rude' as a socially coded meaning" (Williams 2018, p. 159). Finally, in our third category of offenses we have subsumed perturbations in the flow of discourse (vacillations, reformulations, etc.), the use of improper language, the so-called social gaffes (burping, sneezing, etc.) and other incidents of this kind. These cases are likely to produce shame or embarrassment in the speaker himself/herself, but do not imply any offense at all to the interlocutor. Therefore, they occupy the other pole of the continuum.

Interpreting our Mexican data in light of this continuum, we can conclude that Mexican Spanish seems to generalize *perdón* primarily as a self-face-saving device (positive face of speaker), essentially geared towards the face-wants of the speaker but without real offense towards the interlocutor (92%). To a lesser degree, it is also used for minor offenses towards the hearer's negative face (76%), but for the more serious offenses, it competes with the performative verb *disculpar* (43%).

Table 4 compares these tendencies with the Peninsular Spanish corpus data:

Table 4. Relation between (orientation of) *face* and form of apology in Peninsular Spanish.

| Face Affected/Orientation | <i>Disculpar</i> | | <i>Perdonar</i> | | <i>Lo Siento</i> | | <i>Perdón</i> | | Total | |
|---------------------------|------------------|----|-----------------|-----|------------------|-----|---------------|-----|-------|------|
| | # | % | # | % | # | % | # | % | # | % |
| Positive face S | 4 | 2% | 31 | 19% | 6 | 4% | 125 | 75% | 166 | 100% |
| Negative face H | 7 | 4% | 88 | 50% | 15 | 9% | 66 | 37% | 176 | 100% |
| Positive face H | 1 | 1% | 39 | 61% | 14 | 22% | 10 | 16% | 64 | 100% |

The overall distribution of *perdón* stands out: it seems to run along the continuum of seriousness in the same direction as in Mexico, but with some striking diatopic differences. Although both Spanish and Mexican data favor the use of the form *perdón* for offenses towards the positive face of the speaker, Peninsular Spanish does not reach the same level of productivity as Mexican Spanish (75% in Peninsular vs. 92% in Mexican Spanish). Contrary to what happens in the Mexican corpus, in the context of minor offenses towards the negative face of the hearer, Peninsular *perdón* loses its privileged status and competes with its performative counterpart *perdonar* (resp. 37% and 50%). Finally, for the more serious offenses towards the positive face of the interlocutor, *perdón* is very rare (16%). In these cases, the use of *perdonar* also prevails (61%) and—to a lesser extent—*lo siento* is used (22%). In the next section, we will interpret these results against the broader theoretical background of grammaticalization theory and semantic-pragmatic change.

4. Discussion: Diverging Patterns of Grammaticalization

At first glance, our corpus-based study on *perdón* in Mexican and Peninsular Spanish seems to confirm the findings of previous studies in that the wordform *perdón* has undergone a grammaticalization process resulting in multiple discourse-related functions. Indeed, Fuentes Rodríguez (2009) includes *perdón* in her *Diccionario de conectores y operadores del español*, distinguishing between a “modal operator”, used to attenuate the illocutionary force of speech acts threatening to the interlocutor (a request, a refusal, an interruption, etc.), and a “connective reformulation marker”, appealed to in contexts of discourse self-repairs. Similarly, Brenes Peña (2021) organizes the procedural meanings developed by the original apologetic form along three dimensions: metadiscursive (text connective in reformulations, hesitations, repetitions, etc.), interactional (turn-taking system, attention-getter in discourse openings, introducing requests of repetitions and explanations), and argumentative (disagreement marker). Analogous views have been expressed in relation to English *sorry* (see Section 2 above). From the perspective of these authors, the grammaticalization process undergone by *perdón/sorry* is comparable to that of other DMs and developed into a wide array of discourse-related functions that compete with other DMs displaying equivalent functions such as *bueno* (disagreement, dispreferred responses and corrections), *oye/oiga* (to attract attention, or to act as a mitigator in controversial contexts), *o sea* (repair, utterance completion and clarification), *ahora* (disagreement marker), *este* (discourse flow, reformulation, hesitation), amongst others.

However, this competition of different DMs for one and the same function leaves the motivation behind the choices speakers make between available forms unexplained. If we identify *perdón* as a member of a paradigm that comprises other items like *bueno*, *o sea* or *este*, the question arises as to what extent these DMs are interchangeable and what motivates speakers to choose one DM over another to express the “same” function. More specifically, the present study focused on *perdón* in order to answer the question of what makes this marker unique compared with other (apparently) equivalent functional forms. By taking into account the type and seriousness of the offense and the face affected, the present study aspired, then, to both complement and deepen previous findings on this apologetic marker.

Under our proposal, contemporary *perdón* is a polysemous item (see Fischer 2006), basically associated with three pragmatic meanings. The interpretation of the formula

depends on the concrete discourse contexts in which it is used and interacts in a critical way with the type of offense:

- (i) *Perdón*₁ is a genuine expression of regret that is essentially hearer-supportive, but it comes along with a secondary and inferable message of interest to the speaker ('I know the norm; I usually do not offend people');
- (ii) *Perdón*₂ relates to minor social infringements; the illocutionary component of regret is diluted, while the self-protective inferential message gains in prominence;
- (iii) *Perdón*₃ is a grammaticalized self-face-saving device; the remedial move exclusively targets the face-wants of the seemingly apologizing individual. This category also includes the cases related to the flow of discourse.

***Perdón*₁: expression of regret geared towards the hearer**

As explained above (see Section 2), the form *perdón* originates as an ellipsis of the performative verbal expression *pido perdón*, and, up to the present, it can be used as an IFID for expressing genuine apologies. This origin already sets apart *perdón* from other DMs in the same paradigm: it is important to bear in mind that—contrary to what usually happens to other DMs—the point of departure for *perdón* rests in an element that already has an illocutionary force at the onset and this illocutionary material is recruited for further illocutionary and discursive uses (Claridge and Arnovick 2010). Indeed, from a canonical point of view, a speech act of apology is prompted by a wrongdoing—an offense—committed by one person against another, for which the offender takes at least partial responsibility and apologizes to the offended individual in an attempt to repair the damage inflicted on their relationship. In this scenario, the illocutionary meaning of the utterance of apology is an expression of regret (Norrick 1978) on the part of the speaker, who communicates his/her emotional state regarding the situation, and the desired perlocutionary effect of the apology has to do with the hope of being forgiven by the addressee (Norrick 1978; Edmondson 1981). This also explains the common view that apologies are fundamentally hearer-supportive and can therefore be defined as a manifestation of “polite” behavior targeted at the addressee’s face-needs, i.e., attentive to the other’s concern for his/her social image, and intent on boosting the other’s sense of self-worth potentially harmed by the offense (e.g., Owen 1983; Trosborg 1987; Olshtain 1989; Blum-Kulka et al. 1989; Wagner 2004; González-Cruz 2012). In other words, from a canonical point of view, apologies are by nature intersubjective, if this notion is understood as the “expression of Speaker attention to the ‘self’ of addressee [...] in the social sense of paying attention to their ‘face’ or ‘image needs’ associated with social stance and identity” (Traugott 2003, p. 128). Normally, (inter)subjective phenomena arise when lexical items develop pragmatic functions, but expressive speech acts like apologies are interactional by nature and convey attitudes and evaluations of the involved parties as part of their essence.

It is worth noting that apologizers themselves also obtain some benefit as a product of their willingness to engage in repair work. This is suggested in various studies (Holmes 1990; Haverkate 1994; Ogiemann 2006, 2015), and is elaborated upon in Meier (1995, pp. 388–89):

Note that in stark contrast to B/L and those who incorporate their framework, I posit Repair Work to be an image-saving device as regards the Speaker (not the Hearer), making S’s image the central figure. Concern for H’s face is only a by-product of the attempt to save S’s face, “an altruism in egoism” as so aptly puts it. Repair Work is thus an attempt to show that the Speaker is a ‘good guy’ (despite having violated a social norm) and can be relied upon in the future to act predictably in accordance with the social norms of a particular reference group (i.e., to act appropriately). This is a type of reaffirmation of shared values, an uncertainty reduction, which helps to assure S’s membership in the group wherein she or he can derive the same benefits from co-members’ predictable behavior as they can from S’s.

Although we may disagree with Meier’s exclusive emphasis on the speaker-oriented dimension of genuine apologies, the idea that speakers simultaneously pursue subjective goals when they apologize makes sense, considering Goffman’s (1972) hypothesis about both other-directed and self-directed facework moves that occur in social interactions.

As a way of integrating this idea in our definition of genuine (serious) apologies, we propose, drawing on Boye (2023), that an additional, underlying meaning akin to ‘I know the norms of appropriate behavior prevailing in our community and I usually act accordingly’ exists as a pragmatic inference with “discursively secondary status”, which has the potential of becoming conventionalized or grammaticalized in contexts where no real speech act of apology is performed.

This use of *perdón*₁, related to more serious offenses, appears in contexts of disagreements and criticisms (see example 15 above) and also in case of lack of consideration, where the speaker offends the hearer by ignoring, for example, his name (16):

- (16) BAR3M1 [v] Ahora, la tostada. Gloria, solo mermelada sin mantequilla ¿no? CBAR3F6 [v] No no, es Paula. BAR3M1 [v] Ah, Paula *perdón*, me confundió de nombre. (CORMA)
 ‘BAR3M1 [v] Now, the toast. Gloria, just jam with no butter, right? CBAR3F6 [v] No no, it’s Paula. BAR3M1 [v] Ah, Paula *sorry*, I got the wrong name.’

Perdón₂: minor social infringements

Moving further along the continuum of seriousness, consider now the intermediate cases in which people apologize for what seems to be rather minor offenses, primarily motivated by the desire to evince good manners (Norrick 1978). The impression is that, in these circumstances, the words of apology do not express true regret on the part of the speaker—meaning that their original illocutionary force is weakened—serving instead as a “token acknowledgment of some minor infraction or mishap” (Jucker 2019b, p. 20), “associated with the implied message ‘please, don’t think I’m rude’ as a socially coded meaning” (Williams 2018, p. 159). Thus, in comparison with genuine apologies, as defined above and drawing on Boye (2023), we can speak of a shift in the relative prominence of the expression of regret, signaled by *perdón*, and the inferable device of self-protection. What in the case of genuine apologies only occupies “discursively secondary status”—i.e., the speaker’s concern with projecting the image of a person who complies with the rules of socially sanctioned behavior—now comes to occupy center stage, while the remedial work undertaken for the hearer’s sake recedes into the background.

Examples of this category are interruptions in turn-taking contexts and different kinds of impositions on the hearer, as seen above in (14) and again in the following excerpt (17):

- (17) oye pero pero / ¿cómo? / bueno / *perdón* que te regrese al tema / [...] E: eso ya me interesa particularmente / porque creo que me debe interesar // este / pero cómo es / o sea / entonces / ¿el virus del papiloma / ya es cáncer? (PRESEEA)
 ‘hey but / how? / well / *sorry* to get back to the topic / [...] E: I’m particularly interested in this / because I think I should be interested // in this / but how is it / I mean / so / is the papillomavirus / already cancer?’

It is clear that in this kind of examples, the offense rather relates to a minor social infringement, while the self-protective inferential message gains central stage.

Perdón₃: grammaticalized self-face-saving device

The third type of situation in which *perdón* occurs embraces all those cases where one is at a loss trying to determine the “offense” inflicted on the addressee. As mentioned in the previous section, these situations correspond to perturbations in the flow of discourse (vacillations, reformulations, etc.), the use of improper language, the so-called social gaffes (burping, sneezing, etc.) and other incidents of this kind, which are likely to produce a sense of shame or embarrassment in the speaker himself/herself. Various uses of *perdón*₃ were shown in (10) to (13) above. Another case of *lapsus linguae* is illustrated in (18):

- (18) cuando estaba ya, este, trabajando en San Juanito San Juanico, *perdón* dónde varias salchichas de gas estallaron. (CHBC)
 ‘when I was already, huh, working in San Juanito San Juanico, *sorry* where several gas sausages exploded.’

The sense of shame or embarrassment associated with this third type of situations is an experience regarded by psycholinguists as belonging to the class of “self-conscious emotions” and said to arise when individuals evaluate their actions in relation to some standard or rule of acceptable behavior and conclude that they have failed (Lewis et al. 1993). At the same time, this brings to mind the concept of “observed behavior (OB) face processes” recently formulated in Lacroix (2023), according to which speakers engage in self-oriented facework in situations where they think that the addressee, witnessing their improper behavior, will evaluate them negatively, and attempt to counter the potential negative judgement of their interlocutor in some way or another.

We believe that Lacroix’s description fits our cases of no offense to the interlocutor. The self-conscious speaker, aware of his/her failure to comply with the rules of proper—or just adequately articulate—behavior, anticipates how this will affect his/her image in the eyes of the interlocutor and appeals to *perdón* to ward off the potential damage. It is a *perdón* stripped of its original illocutionary meaning (‘I regret that I offended you’). The only message it carries is something along the lines of ‘I know how I am expected to behave; my action should be seen as an unwonted slip’. In other words, the pragmatic inference, available but subordinated in genuine apologies (*perdón*₁), has been incorporated into the semantics of *perdón* and has given rise to a conventionalized or grammaticalized meaning which enables *perdón* to function as a self-face-saving device (*perdón*₃). Of course, *perdón*₂ and *perdón*₃ are proximate: they share the prominence of the self-oriented value but differ in that *perdón*₂ still contains a (weak) expression of regret for an acknowledged offense caused to the addressee.

From this point of view, the three meanings of *perdón* are susceptible to being ordered along a cline of decreasing intersubjectivity and increasing subjectivity. This goes against the traditional direction posited by Traugott (1999, p. 3) according to which intersubjectification follows, and arises from, subjectification, and which has been verified in a wide range of concrete cases to account for the evolution of DMs. However, this unidirectional shift from subjectivization to intersubjectivization has also been challenged in some studies, suggesting that the relation between the two notions should be thought as allowing for variable patterns of development (cf. among others Cornillie 2014; Hancil 2018 and references therein).

This in-depth analysis of *perdón* also enables us to tackle the question of what exactly differentiates *perdón*₃ from other DM like *bueno*, *oye/oiga*, *este*, etc. connected with similar metadiscursive functions (reformulations, hesitations, repetitions, etc.) and to pinpoint more precisely its specific contribution to this paradigm of DMs. Based on our empirical analysis, it is clear that *perdón*₃ addresses problems related to the flow of discourse in its own unique way, focusing the perspective on the image of self. The competing forms have different histories and introduce different nuances in the management of these conversational phenomena.

In a similar vein, *perdón*₂ invites an analysis in terms of an interactional type of connective marker (Brenes Peña 2021) or modal operator (Fuentes Rodríguez 2009) that, considering the functions it performs in contexts of minor offenses, is evaluated to be so slight that *perdón* is said to express a “pseudo-apology” (Brenes Peña 2021, pp. 156–57). In these cases, it is used to soften the impoliteness of interruptions, of requests for repetitions, of unwelcome responses to a petition, or of intrusions into one’s territory with an attention-getting marker. Again, some of these functions have been attributed to other DMs (*bueno*, *oye/oiga*, *mira/mire*), and the relevant question hinges on what it is that *perdón*₂ accomplishes in these contexts, in contrast with the competing forms. We have grouped these contexts in our category of offensive behavior threatening the hearer’s negative face, since they involve violations of one’s claims to privacy and freedom from impositions and

impediments (in terms of Brown and Levinson 1987). We consider that, if speakers choose *perdón* over other markers in situations of this nature, it is because they feel that some words of apology are in order. However, sentiments of genuine regret and the hope of being forgiven are clearly absent. The main preoccupation concerns the negative impact the offense, however slight, will have on their public image. This balance between acknowledging the offense and attending interests of the self is precisely what *perdón*₂ helps to achieve, with its backgrounded apologizing value and its foregrounded message of face-redress ('I know the social rules and I usually respect them').

Interestingly, from a contrastive point of view, in our Mexican data, where *perdón* predominates, disagreements and criticisms, classified as more serious offenses on the grounds that they challenge the other's social dignity (his positive face-wants), still resist the use of *perdón* to some extent. We interpret this phenomenon as suggestive of the fact that the entrenchment of grammaticalized *perdón*₃, along with the expansion of the proximate *perdón*₂, have generated an implicit association of the form with issues of self-worth such that speakers hesitate to resort to the formula (*perdón*₁) in contexts where they evaluate their behavior as being truly offensive and choose more elaborate expressions of apology, such as *disculpa/e*, to convey their feeling of regret.

The Peninsular data, on the other hand, give evidence of a less advanced process of grammaticalization. The self-face-saving device (*perdón*₃) is frequent in contexts of no offense (to the interlocutor), but elsewhere, other IFIDs (especially *perdona/e*, *lo siento*), are still preferred. This suggests that, contrary to Peninsular Spanish, Mexico seems to have also regularized the use of *perdón* for offenses harming the negative face of the addressee, with a clear dominance in turn-taking contexts. This divergence between peninsular and Mexican *perdón* suggests a further stage of the latter in its grammaticalization towards discourse organization. This synchronic variation (Schneider and Barron 2008; Aijmer 2022) corroborates the well-known fact that the use of pragmatic markers not only differs between languages, but also between varieties of the same language and thus reveals diverging patterns of ongoing grammaticalization between both varieties.

5. Concluding Remarks

By means of a corpus-based comparative analysis, this study has examined the degree of grammaticalization of the apologetic marker *perdón* in Peninsular and Mexican Spanish, which has led to a number of significant insights situated at both the methodological and theoretical level of analysis.

First of all, from a methodological point of view, we approached the speech act of apology from the perspective of the two basic theoretical concepts inherent to an apology, namely the type of offense and the concept of face. The analytical tool used for this purpose is a taxonomy of offenses motivating apologies in Spanish, organized around the concept of face of both the speaker and the interlocutor. This taxonomy is shown to be a fruitful methodological tool for the analysis of apologetic markers that provides a systematic and verifiable alternative to more intuitive approaches. It does not only offer usage-based evidence for previous theorizing concerning the grammaticalization process of apologetic markers, but also leads to a gradual refinement of these previous results from a contrastive point of view.

Theoretically, the study offers a comprehensive perspective on the grammaticalization of *perdón* in Spanish. Contrary to what usually happens to other DMs, the point of departure for *perdón* rests in an element that already has an illocutionary force at the onset and this illocutionary material is recruited for further illocutionary and discursive uses. That is, the grammaticalization or discursification process of *perdón* embodies a process that affects the illocutionary potential of the speech act itself. So, rather than semantic bleaching, this can be best described as a case of progressive weakening of its illocutionary force (Jucker 2019a). However, even in this grammaticalized use of the form, some important vestiges of the canonical model of apologies remain essential for the comprehension of its present-day uses: we have seen that the two theoretical notions underlying the canonical

definition of an apology—namely face and offense—are still determining the contemporary grammaticalized uses and values of *perdón*. At the same time, these vestiges also help to relate its contemporary grammaticalized values to the original illocutionary act of regret. More concretely, under our proposal, contemporary *perdón* is a polysemous item associated with three main pragmatic meanings. The interpretation of the formula depends on the concrete discourse contexts in which it is used and interacts in a critical way with the type of offense determining its appearance:

- *Perdón*₁ is an expression of regret that is essentially hearer-supportive, but it comes along with a secondary and inferable message of interest to the speaker ('I know the norms; I usually do not offend people');
- *Perdón*₂ relates to minor social infringements; the illocutionary component of regret is diluted, while the self-protective inferential message gains in prominence;
- *Perdón*₃ is a grammaticalized self-face-saving device; the remedial move exclusively targets the face-wants of the seemingly apologizing individual.

That is to say, the bleaching of the original illocutionary force of the speech act of apology geared towards the interlocutor is compensated by a pragmatic strengthening of what used to be only a pragmatic inference with discursively secondary status in the case of genuine apologies. More precisely, the underlying meaning akin to 'I know the norms of appropriate behavior prevailing in our community and I usually act accordingly' has become conventionalized as a grammaticalized formulaic speech act. In this grammaticalized form, it is often used without the presence of a clear offense towards the interlocutor but converts into a self-face-saving device geared towards the speaker. In other words, the pragmatic inference, available but subordinated in genuine apologies (*perdon*₁), has been incorporated into the semantics of *perdón* and has given rise to a conventionalized or grammaticalized meaning which enables *perdón* to function as a self-face-saving device (*perdón*₃).

Interestingly, the degree of entrenchment of grammaticalized *perdón*₃ seems to give rise to diverging patterns of grammaticalization across varieties of the same language. As such, we have seen that Mexican Spanish seems to have regularized the use of *perdón* not only as a self-face-saving device (*perdón*₃), but also frequently allows it for offenses harming the negative face of the addressee, with a clear dominance in turn-taking contexts. Peninsular Spanish, on the other hand, gives evidence of a less advanced process of grammaticalization. The self-face-saving device (*perdón*₃) is frequent in contexts of no offense (to the interlocutor), but, elsewhere, other IFIDs (especially *perdona/e, lo siento*) are still preferred. To conclude, the present study thus corroborates and at the same time refines Viberg's (1999) conclusion that grammaticalization can drive cognates apart semantically, as long as we interpret cognates both at the interlinguistic and intralinguistic/dialectic level.

Author Contributions: Conceptualization, M.J. and C.M.; Methodology, C.M. and J.E.A.B.; Investigation, M.J., C.M. and J.E.A.B.; Data curation, M.J. and J.E.A.B.; Writing—original draft, M.J. and C.M.; Writing—review & editing, M.J.; Visualization, M.J.; Supervision, M.J.; Project administration, M.J.; Funding acquisition, M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ghent University, grant number [BOF/STA/202009/031].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data of this study is available in TROLLING. Jansegers, Marlies; Melis, Chantal; Arrington Báez, Jennie Elenor, 2023, Replication Data for: Diverging grammaticalization patterns across Spanish varieties: the case of *perdón* in Mexican and Peninsular Spanish, <https://doi.org/10.18710/IEXVVN> (accessed on 1 December 2023), DataverseNO.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ See Section 3.1 for more information related to the corpora used for this study.
- ² Examples taken from González-Cruz (2012).
- ³ Although theoretically possible, our dataset does not contain any cases of offenses towards the speaker's negative face.

References

Corpus Source

- América y España español coloquial* [AMERESCO] [online]. <https://esvaratenuacion.es/> (accessed on 1 December 2023).
- Valencia Español Coloquial* [Val.Es.Co] [online]. <https://www.valesco.es/> (accessed on 1 December 2023).
- Corpus del Habla de Baja California* [CHBC] [online]. <http://www.corpus.unam.mx/geco/portal/index/chbc> (accessed on 1 December 2023).
- Corpus Michoacano del Español* [CME]. Non published corpus, access courtesy of the authors.
- Corpus Oral de Madrid* [CORMA]. Non published corpus, access courtesy of the authors.
- Corpus Oral de Referencia del Español Contemporáneo* [CORLEC] [online]. <http://www.llf.uam.es/ESP/Corlec.html> (accessed on 1 December 2023).
- Corpus Sociolingüístico de la Ciudad de México* [CSCM] [online]. https://lef.colmex.mx/corpus_sociolingüístico.html (accessed on 1 December 2023).
- Proyecto para el Estudio Sociolingüístico del Español de España y América* [PRESEEA] [online]. <https://presea.uah.es/> (accessed on 1 December 2023).

Secondary Source

- Aijmer, Karin. 1996. *Conversational Routines in English: Convention and Creativity*. Studies in Language and Linguistics. London: Longman.
- Aijmer, Karin. 2022. "Well He's Sick Anyway Like": *Anyway* in Irish English. *Corpus Pragmatics* 6: 101125. [CrossRef]
- Arnovick, Leslie K. 1999. *Diachronic Pragmatics: Seven Case Studies in English Illocutionary Development*. Amsterdam and Philadelphia: John Benjamins.
- Bello, Bethany Ann. 2015. Atenuación e intensificación pragmáticas en la expresión de actividades de imagen: Un estudio contrastivo en conversaciones coloquiales del español y del inglés. Ph.D. thesis, University of Valencia, Valencia, Spain.
- Blum-Kulka, Shoshana, Juliane House, and Gabriele Kasper. 1989. *Cross-Cultural Pragmatics: Requests and Apologies*. Norwood: Ablex.
- Boye, Kasper. 2023. Grammaticalization as Conventionalization of Discursively Secondary Status: Deconstructing the Lexical-Grammatical Continuum. *Transactions of the Philological Society* 121: 270–92. [CrossRef]
- Bravo, Diana. 1999. ¿Imagen positiva vs. imagen negativa? Pragmática socio-cultural y componentes de face. *Oralia* 2: 155–84.
- Bravo, Diana. 2005. Categorías, Tipologías y Aplicaciones. Hacia una Redefinición de la Cortesía Comunicativa. In *Estudios de la (Des)Cortesía en Español. Categorías Conceptuales y Aplicaciones a Corpora Orales y Escritos*. Edited by Diana Bravo. Buenos Aires: Dunken, pp. 21–52.
- Brenes Peña, Ester. 2021. De la petición de disculpas a la contraargumentación: Análisis macrosintáctico de perdón. In *Sintaxis Discursiva: Construcciones y Operadores en Español*. Edited by Catalina Fuentes Rodríguez, Ester Brenes Peña and Víctor Pérez Béjar. Bern: Peter Lang, pp. 135–71.
- Brinton, Laurel J. 2017. *The Evolution of Pragmatic Markers in English: Pathways of Change*. Cambridge: Cambridge University Press. [CrossRef]
- Brooks, Roy L. 1999. *When Sorry Isn't Enough: The Controversy over Apologies and Reparations for Human Injustice*. New York: New York University Press.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Castillo Lluch, Mónica. 2008. La formación de marcadores discursivos vaya, venga, anda y vamos. In *Actas del VII congreso internacional de historia de la lengua española*. Edited by Concepción Company Company. Madrid: Arco Libros, pp. 1739–52.
- Chen, Rong. 2001. Self-Politeness: A Proposal. *Journal of Pragmatics* 33: 87–106. [CrossRef]
- Claridge, Claudia, and Leslie Arnovick. 2010. Pragmaticalisation and Discursification. In *Historical Pragmatics*. Edited by Andreas H. Jucker and Irma Taavitsainen. Berlin and New York: De Gruyter Mouton, pp. 165–92.
- Cornillie, Bert. 2014. La historia de la complementación con parecer y resultar. Apuntes sobre la (inter)subjetivización. *Revista de la Sociedad Argentina de Lingüística* 1: 1–15.
- Coulmas, Florian. 1981. Poison to Your Soul: Thanks and Apologies Contrastively Viewed. In *Conversational Routine*. Edited by Florian Coulmas. The Hague: Mouton, pp. 69–91.
- Denoyelle, Corinne. 2020. La réalisation de l'excuse en moyen français: Une recherche en pragmatique historique. *Travaux de Linguistique* 2: 145–81. [CrossRef]
- Detges, Ulrich, and Richard Waltereit. 2016. Grammaticalization and pragmaticalization. In *Manual of Grammatical Interfaces in Romance*. Edited by Susann Fischer and Christoph Gabriel. Berlin and Boston: De Gruyter, pp. 635–58.
- Deutschmann, Mats. 2003. Apologizing in British English. Ph.D. thesis, University of Umeå, Umeå, Sweden.

- Edmondson, Willis J. 1981. On Saying You're Sorry. In *Conversational Routine*. Edited by Florian Coulmas. The Hague: Mouton, pp. 273–88.
- Fischer, Kerstin. 2006. *Approaches to Discourse Particles*. Amsterdam: Elsevier.
- Fuentes Rodríguez, Catalina. 2009. *Diccionario de Conectores y Operadores del Español*. Madrid: Arco Libros.
- Gancedo Ruiz, Marta. 2019. Evolución de la imagen del rol familiar en el teatro de finales del siglo XIX a mitad del XX. Su manifestación en la atenuación e intensificación de los actos directivos. Ph.D. thesis, University of Valencia, Valencia, Spain.
- Ghezzi, Chiara, and Piera Molinelli. 2019. Italian *scusa* from politeness to mock politeness. *Journal of Pragmatics* 142: 245–57. [CrossRef]
- Gibney, Mark, Rhonda Howard-Hassman, Jean-Marc Coicaud, and Nicklaus Steiner. 2007. *The Age of Apology: The West Faces Its Own Past*. Philadelphia: University of Pennsylvania Press.
- Goffman, Erving. 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Doubleday-Anchor.
- Goffman, Erving. 1971. *Relations in Public: Microstudies of the Public Order*. London: Allen Lane.
- Goffman, Erving. 1972. On Face-Work: An Analysis of Ritual Elements in Social Interaction. In *Communication in Face-to-Face Interaction*. Edited by John Laver and Sandy Hutcheson. Harmondsworth: Penguin, pp. 179–96.
- González-Cruz, María-Isabel. 2012. Apologizing in Spanish: A study of the strategies used by university students in Las Palmas de Gran Canaria. *Pragmatics* 22: 543–65. [CrossRef]
- Hancil, Sylvie. 2018. (Inter)subjectification and paradigmaticization: The case study of the final particle but. In *New Trends in Grammaticalization and Language Change*. Edited by Sylvie Hancil, Tine Breban and José Vicente Lozano. Amsterdam: John Benjamins, pp. 291–313.
- Harris, Sandra, Karen Grainger, and Louise Mullany. 2006. The Pragmatics of Political Apologies. *Discourse and Society* 17: 715–37. [CrossRef]
- Haverkate, Henk. 1994. *La cortesía verbal: Estudio pragmatolingüístico*. Madrid: Gredos.
- Hernández Flores, Nieves. 2005. La cortesía como búsqueda del equilibrio de la imagen social: La oscilación de la imagen en un debate televisivo. In *Actas del II Coloquio del Programa EDICE*. Edited by Jorge Murillo Medrano. San José: Universidad de Costa Rica, pp. 37–52.
- Hernández Flores, Nieves. 2008. Politeness and other types of facework: Communicative and social meaning in a television panel discussion. *Pragmatics* 18: 577–603. [CrossRef]
- Holmes, Janet. 1990. Apologies in New Zealand English. *Language in Society* 19: 155–99. [CrossRef]
- Holmes, Janet. 1998. Apologies in New Zealand English. In *The Sociolinguistics Reader: Gender and Discourse*. Edited by Jenny Cheshire and Peter Trudgill. London: Arnold, vol. 2, pp. 201–39.
- Hopper, Paul J. 1991. On some principles of grammaticization. In *Approaches to Grammaticalization*. Edited by Elizabeth Closs Traugott and Bernd Heine. Amsterdam: John Benjamins, vol. 1, pp. 17–36.
- Jucker, Andreas H. 2019a. Speech Act Attenuation in the History of English: The Case of Apologies. *Glossa: A Journal of General Linguistics* 4: 1–25. [CrossRef]
- Jucker, Andreas H. 2019b. “Oops, I forgot, sorry”: The spill cries oops and whoops in the history of American English. *Lingue e Linguaggi* 31: 15–33.
- Kerbrat-Orecchioni, Catherine. 1992. *Les Interactions Verbales*. Paris: Armand Colin, vol. 2.
- Kohnen, Thomas. 2017. Non-Canonical Speech Acts in the History of English. *Zeitschrift für Anglistik und Amerikanistik* 65: 303–18. [CrossRef]
- Lacroix, René. 2023. Two phenomena behind the terminology of face. *Journal of Politeness Research* 19: 323–53. [CrossRef]
- Lazare, Aaron. 2004. *On Apology*. Oxford: Oxford University Press.
- Leech, Geoffrey Neil. 1983. *Principles of Pragmatics*. London: Longman.
- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*, 3rd ed. Berlin: Language Science Press. First published 1995.
- Lewis, Michael, Jeannette M. Haviland-Jones, and Lisa Feldman Barrett, eds. 1993. *The Handbook of Emotion*. New York: Guilford Publications.
- Márquez-Reiter, Rosina. 2000. *Linguistic Politeness in Britain and Uruguay: A Contrastive Study of Requests and Apologies*. Amsterdam: John Benjamins.
- Medina López, Javier. 2023. Formas de perdón, arrepentimiento y disculpas en la historia del español: Una aproximación desde el análisis textual. *Nueva Revista de Filología Hispánica* 71: 499–529. [CrossRef]
- Meier, Ardith J. 1995. Passages of politeness. *Journal of Pragmatics* 24: 381–92. [CrossRef]
- Molina, Clara. 2011. Routes for development in the pragmaticalization of *sorry* as a formulaic marker. *Revista Alicantina de Estudios Ingleses* 24: 191–212. [CrossRef]
- Norrick, Neal R. 1978. Expressive Illocutionary Acts. *Journal of Pragmatics* 2: 277–91. [CrossRef]
- Norrick, Neal R. 1979. The lexicalization of pragmatic functions. *Linguistics* 17: 671–86. [CrossRef]
- Ogiermann, Eva. 2006. Cultural variability within Brown and Levinson's politeness theory. English, Polish and Russian apologies. In *Studies in Contrastive Linguistics*. Edited by Cristina Mourón-Figueroa and Teresa Moralejo-Gárate. Santiago de Compostela: Universidade de Santiago de Compostela Publicacións, pp. 707–18.
- Ogiermann, Eva. 2009. *On Apologising in Negative and Positive Politeness Cultures*. Amsterdam and Philadelphia: John Benjamins.
- Ogiermann, Eva. 2015. Apology discourse. In *The International Encyclopedia of Language and Social Interaction*. Edited by Karen Tracy. Chichester: John Wiley & Sons, pp. 1–16. [CrossRef]

- Olshtain, Elite. 1989. Apologies Across Languages. In *Cross-Cultural Pragmatics: Requests and Apologies*. Edited by Shoshana Blum-Kulka, Juliane House and Gabriele Kasper. Cambridge: Cambridge University Press, pp. 155–73.
- Olshtain, Elite, and Andrew Cohen. 1983. Apology: A Speech Act Set. In *Sociolinguistics and Language Acquisition*. Edited by Nessa Wolfson and Elliot Judd. Rowley: Newbury House, pp. 18–36.
- Owen, Marion. 1983. *Apologies and Remedial Interchanges. A Study of Language Use in Social Interaction*. Berlin: Mouton.
- Schneider, Klaus Peter, and Anne Barron, eds. 2008. *Variational Pragmatics. A Focus on Regional Varieties in Pluricentric Languages*. Amsterdam: John Benjamins.
- Searle, John R. 1969. *Speech Acts*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth C. 1999. From subjectification to intersubjectification. Paper presented at the Workshop of Historical Pragmatics, Fourteenth International Conference on Historical Linguistics, Vancouver, BC, Canada, August 9–13.
- Traugott, Elizabeth C. 2003. Constructions in grammaticalization. In *The Handbook of Historical Linguistics*. Edited by Brian D. Joseph and Richard D. Janda. Oxford: Blackwell, pp. 624–47.
- Trosborg, Anna. 1987. Apology strategies in natives/nonnatives. *Journal of Pragmatics* 11: 147–67. [CrossRef]
- Trosborg, Anna. 1995. *Interlanguage Pragmatics: Requests, Complaints and Apologies*. Berlin and New York: De Gruyter Mouton.
- Van Bogaert, Julie. 2011. “I Think” and Other Complement-Taking Mental Predicates: A Case of and for Constructional Grammaticalization. *Linguistics* 49: 295–332. [CrossRef]
- Viberg, Ake. 1999. The polysemous cognates Swedish *gå* and English *go*: Universal and language-specific characteristics. *Languages in Contrast* 2: 87–113. [CrossRef]
- Wagner, Lisa C. 1999. Towards a Sociopragmatic Characterization of Apologies in Mexican Spanish. Ph.D. thesis, The Ohio State University, Columbus, OH, USA.
- Wagner, Lisa. C. 2004. Positive- and negative-politeness strategies: Apologizing in the speech community of Cuernavaca, Mexico. *International Communication Studies* 13: 19–27.
- Wichmann, Anne, Anne-Marie Simon-Vandenberg, and Karin Aijmer. 2010. How prosody reflects semantic change: A synchronic case study of of course. In *Subjectification, Intersubjectification and Grammaticalization*. Edited by Kristin Davidse, Lieven Vandelandotte and Hubert Cuyckens. Berlin and New York: De Gruyter Mouton, pp. 103–54.
- Williams, Graham. 2018. *Sincerity in Medieval English Language and Literature*. London: Palgrave Macmillan.
- Winter-Froemel, Esme. 2014. Re(de)fining grammaticalization from a usage-based perspective Discursive ambiguity in innovation scenarios. *Folia Linguistica* 48: 503–56. [CrossRef]
- Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. New York: Hafner Publishing Company.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

The Grammaticalization of the Discourse Marker *genre* in Swiss French

Delin Deng

Department of Linguistics, University of Florida, Gainesville, FL 32611, USA; ddeng@ufl.edu

Abstract: By conducting an apparent-time analysis of the OFROM corpus collected in Francophone Switzerland, this study examined the use of *genre* as discourse marker in the speech of 306 French L1 speakers. First, we examined the effect of extralinguistic factors on the discursive use of *genre*. The logistic mixed-effects regression analysis results revealed that the emerging use of *genre* is indeed an ongoing change led by female speakers in Swiss French. This use was favored by monolinguals in Francophone Swiss. Second, we examined the vowel reduction of the DM *genre* in the corpus. Our results revealed that speakers who received only a high school education favor the vowel reduction in the DM *genre* the most. Given the high percentage of phonological reduction in the DM *genre*, we believe that the grammaticalization of this particle has reached its advanced stage in Swiss French. Compared to previous findings on the emerging use of *genre* in Hexagonal French, we suggested that the grammaticalization of the particle *genre* in Swiss French may be independent of that in Hexagonal French. The grammaticalization in Swiss French was much more advanced than in Hexagonal French. This study supplied comparable results on the grammaticalization of the same particle in two different Francophone countries.

Keywords: DM; grammaticalization; phonological reduction; variation and change

Citation: Deng, Delin. 2023. The Grammaticalization of the Discourse Marker *genre* in Swiss French.

Languages 8: 28. <https://doi.org/10.3390/languages8010028>

Academic Editors: Renata Enghels, Marlies Jansegers and Tom Bossuyt

Received: 15 November 2022

Revised: 9 January 2023

Accepted: 11 January 2023

Published: 16 January 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Different scholars have documented the emerging discursive use of *genre* in native French (see, for example, Denison 2002; Mihatsch 2010; Secova 2011). Secova (2011), in particular, reported the use of *genre* as a discourse marker (DM) to be typical in youth language. Examples (1) and (2) illustrate the DM and non-DM use of *genre* in native speech.

- (1) le langage que j'ai des fois *genre* avec mes copines (OFROM_unine19-002)
"the language I sometimes have like with my girlfriends"
- (2) enfin ouais plus ce *genre* de mots comme spoiler (OFROM_unine19-017)
"well yes more this kind of words like spoiler"

As shown by these two examples, when used as a DM, a lexical item no longer belongs to the original category assigned to it. The original use of *genre* is a noun, as in example (2), while the DM *genre* in example (1) functions more like an adverb. The difference between this discursive and non-discursive use is that the removal of the former does not affect the semantic integrity of the sentence, while removing the latter does. This change from non-discursive use to discursive use of a lexical item often involves the process of grammaticalization.

Over recent decades, much research has been dedicated to describing the discursive functions of *genre*. Not much has been known regarding its use in a social context. Only marginal studies discussed this aspect by pointing out that this might be a feature of youth speech (see, for example, Secova 2011). However, we still do not know, for example, if the DM *genre* is fully grammaticalized in Swiss French or if it is still a change in progress. What is the supporting evidence for its grammaticalization if it is a change in progress? Is it a

change independent of that in Hexagonal French? What are the constraining social factors influencing this discursive use? etc.

Therefore, by conducting an apparent-time variationist analysis of the particle *genre* in Swiss French native speech, we hope to be able to answer, if not all, at least some of the questions raised here. The objective of the current work is to, from a quantitative point of view, on the one hand, discuss the grammaticalization of *genre* in Swiss French native speech by analyzing oral data taken from the online publicly accessible corpus and its correlation with social factors, such as age, gender, sociolinguistic situation, and socio-educational status of the speakers. On the other hand, it will also provide supporting evidence of a phonological reduction in the DM *genre* for its grammaticalization.

The structure of this article is as follows: Section 2, the relevant literature on grammaticalization and discursive functions of *genre* in French native speech is reviewed. Section 3, the methodology, including information on corpora, speakers, tokens and extralinguistic factors, as well as statistical analysis, is presented. Section 4, the results of the current study are presented and discussed in tables. Section 5, the article is concluded with a summary of the current work as well as future implications.

2. Theoretical Backgrounds

2.1. Process of Grammaticalization

Over the past century, ample research and debates have contributed to identifying the different stages of grammaticalization to explain how a lexical item becomes grammaticalized. It is believed that the notion of “grammaticalization” was first mentioned by Meillet (1912). He described grammaticalization as “the passage of an autonomous word into the role of grammatical element” (Meillet 1912, p. 131). Heine and Reh (1984, p. 15) argued that grammaticalization is “an evolution whereby linguistic units lose in semantic complexity, pragmatic significance, syntactic freedom, and phonetic substance”. Traugott (1995, p. 1) defined grammaticalization as “the process whereby lexical material in highly constrained pragmatic and morphosyntactic contexts becomes grammatical, in other words, that lexical material in specifiable syntactic functions comes to participate in the structural texture of the language, especially its morphosyntactic constructions”.

It is argued that, according to many, the process of grammaticalization is “unidirectional” (see, for example, Meillet 1912; Saxena 1995; Vincent 2001) in the sense that this process is irreversible (see, for example, Haspelmath 1999). Heine et al. (1991) believed that grammaticalization involves a combination of discrete stages and continuum. As pointed out by Kuryłowicz (1965, p. 52), “grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status”. “Such a progression often involves a number of intermediate stages, making it difficult in many cases to maintain a neat distinction between lexical and grammatical elements” (Saxena 1995, p. 352).

Lehmann identified three parameters and processes of grammaticalization (Lehmann [1982] 1995, p. 164), as illustrated in Table 1:

Table 1. Lehmann [1982] (1995, p. 164) Correlation of grammaticalization process (excerpt).

| Parameter | Weak GR | Process | Strong GR |
|-------------------------|---|--------------|--|
| Scope | Item relates to constituent of arbitrary complexity | Condensation | item modifies word or stem |
| Bondedness | Item independently juxtaposed | Coalescence | item = affix, phonol, feature or carrier |
| Syntagmatic variability | Item can be shifted around freely | Fixation | item occupies fixed slot |

Within this framework, grammaticalization is viewed as a process of increased bondedness but decreased scope. Two aspects are central to the process of grammaticalization: phonological reduction and semantic bleaching. While some believe that phonological reduction and semantic bleaching often go hand in hand in the process of grammaticaliza-

tion, Haspelmath (1999, p. 1058) argued that semantic bleaching should be the cause of other processes of grammaticalization in that “lexical items that fulfill a frequent discourse function will then increase in frequency because they are very often useful”. He further pointed out that “increased frequency also means increased predictability, and the more predictable an item is, the less phonologically salient it needs to be” (Haspelmath 1999, p. 1058). This is mainly because when it becomes highly frequent, the probability of misunderstanding the lexical item with phonological reduction decreases accordingly.

Later, Lehmann (2015) further identified different phases of grammaticalization.

As illustrated in Figure 1, the morpho-phonemic change only occurs at a later phase of grammaticalization. As pointed out by Haspelmath (1999), it is at this stage of development that the lexical item in question has reached a high frequency in use and that the phonological reduction does not affect the listener’s comprehension. That is to say, if we can identify any phonological reduction in the grammaticalization of a lexical item, the process of grammaticalization for this lexical item has probably reached a relatively advanced stage.

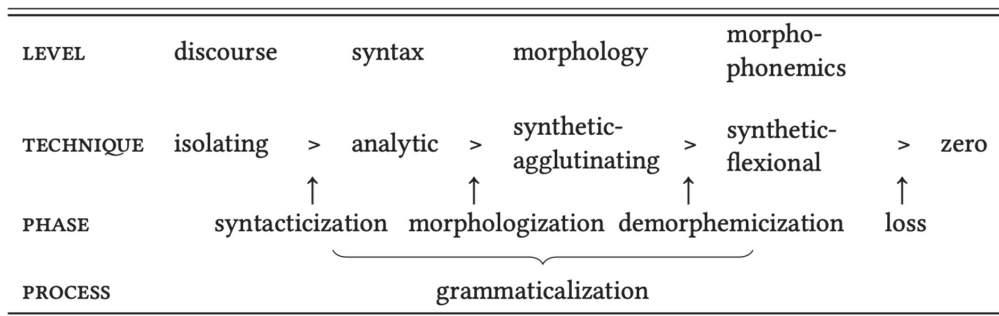


Figure 1. Phases of grammaticalization (Lehmann 2015, p. 15).

This discussion of grammaticalization becomes even more relevant when talking about DMs. As defined by Hansen (1998), DMs are “linguistic items which fulfill a non-propositional, meta-discursive (primarily connective) function, and whose scope is inherently variable, such that it may comprise both sub-sentential and supra-sentential units”. It is usually the case that when a lexical item is used as a DM, it no longer belongs to the original grammatical category assigned to it but is semantically independent of the sentence. Removing a DM from the carrying sentence does not affect the grammaticality of the sentence. The impact is only at the pragmatic level. As pointed out by Fraser (1988, p. 22), “the absence of the DM does not render a sentence ungrammatical and/or unintelligible. It does, however, remove a powerful clue about what commitment the speaker makes regarding the relationship between the current utterance and the prior discourse”.

2.2. Genre Used as a DM in Native Speech

When did *genre* start to be used as a DM? What are the pragmatic functions it fulfills when used as a DM? Mihatsch (2010) pointed out that the form of *genre de* began to be used as an approximator as early as the 15th century. Nevertheless, it is generally believed that *genre*, the shortened form of *genre de*, is a newly emerged and very frequent DM in European French. As confirmed by Secova (2011), it is difficult to establish the diachronic development of the appearance of *genre* as a particle since it is not documented in any dictionaries and only exists in some spoken corpora. Denison (2002) assumed that the discursive use of *genre* probably developed from its qualifying particle use by semantic bleaching and blurring of scope boundaries. It is noticed that the English *like*, Canadian French *comme*, and European French *genre* do share some pragmatic similarities. Mihatsch (2010) identified three functions of *genre* in European French: adaptor, quotative use, and rounder. Secova (2011) differentiated between approximation; exemplification/paraphrasing; reporting

speech, thought and attitude; and expressing irony. She also pointed out that at the epistemic level *genre* seems to be used as a hedge while at the syntax-semantic level as a marker of focus. We will illustrate with examples below based on Secova's classification. The same functions could be identified in both Hexagonal and Swiss French.

2.2.1. Approximation

It is the primary function of *genre* as a particle to indicate the approximation or inexactitude, functioning both as adaptors and rounders.

- (3) ils ont fait du bruit franchement . . . il était un truc *genre* trois—quatre heures du matin
(Secova 2011, p. 100)
"they made noises frankly . . . it was a thing like three-four o'clock in the morning"

It is clear here that *genre* modifies the numeral value and overlaps the rounder function of *comme* in example (3). *Genre* could be paraphrased by *environ* or *à peu près*. By using *genre*, the speaker indicates the vagueness of what was said by attenuating the force of the speech act.

2.2.2. Exemplification

In this case, *genre* is often used in the utterance-initial position and connects the previous utterance with the following utterance. *Genre* either provides an example or a justification for the previous utterance.

- (4) elle parle toute seule . . . *genre* t'as vu quand elle était sur le canapé? (Secova 2011, p. 104)
"she talks all alone . . . like you saw when she was on the couch?"

2.2.3. Speech Reporting Speech

The quotative *genre* can usually be substituted by the verb *say*. It introduces direct speech without being too serious about what is quoted. In this sense, the use of *genre* reduces the formality of the speech.

- (5) moi j'ai bien aimé . . . franchement . . . mais Patrick il était là *genre* "ouais j'aime pas la chanteuse" (Secova 2011, p. 106)
me I loved a lot . . . frankly . . . but Patrick he was there like "yeah I don't like the singer"

2.2.4. Expressing Irony

It seems that this function is unique to *genre*. The ironic *genre* often offers a quotation or explanation with an ironic tone and could probably be paraphrased by the French phrase *soi-disant* "so-called". By using *genre*, the speaker reduces the credibility of the utterance.

- (6) tu sais à quelle heure elle nous remplace son cours *genre* pour pas nous déranger? à huit heures Samedi!!! (Fleischman and Yaguello 2004, p. 137)
"you know at what time she replaces us with her lesson like not to disturb us? at eight o'clock Saturday!!!"

As shown in example (6), the fact that the replacement of the class is at eight o'clock on a Saturday morning is an act of disturbing others. After the ironic *genre*, the speaker used the expression "not to disturb us," where he actually meant the opposite. The *genre* here constitutes a drastic contrast between what is said and what is intended by the speaker. In some way, *genre* reinforces the force of the hidden meaning.

Despite the depth of research on discursive functions of *genre* and the historical development of its use as a DM, to the best of our knowledge, no research has documented the phonological reduction of this particle when used as a DM. However, as illustrated in the literature, phonological reduction is strong evidence for the grammaticalization of lexical items. Previous studies have mainly focused on describing its discursive functions. There has not been any attempt to explore how social factors, such as the age, gender, or socio-economic status of the speakers, could impact its use as a DM. Previous studies on language variation and change have demonstrated that social factors could be important

indicators for the development of linguistic variables in a speech community (see, for example, Labov 2001). A variationist study of *genre* at binary classification of DM *genre* and non-DM *genre* could shed more light on this topic.

Therefore, in this work, by conducting logistic mixed-effects regression analysis in R using Rbrul (Johnson 2009; R Core Team 2021) on *genre* in OFROM corpus (le corpus Oral de Français de Suisse Romande, (Avanzi et al. 2012–2020), www.unine.ch/ofrom (accessed on 11 November 2022), we tried to answer the following research questions: Is this discursive use of *genre* an ongoing change in Swiss French in apparent time? What are the social factors that impact this discursive use? Is there any other evidence, such as phonological reduction, that supports the claim that *genre* is undergoing the process of grammaticalization?

3. Methodology

3.1. Corpus

The corpus used in this study is OFROM. OFROM is a text-sound-aligned publicly accessible online corpus consisting of some sociolinguistic interviews conducted with speakers of French in Switzerland. All the interviews were orthographically transcribed in Praat (Boersma and Weenink 2021). The corpus was initiated in 2012 and is still under construction. As of 2020, the corpus contains 342 speakers, totaling 64 h of recording and 1,005,000 words. For the current study, we only used data collected in seven Francophone cantons, where French is the official language, in Switzerland. Therefore, only 306 speakers were included in our final analysis. These seven cantons are Neuchâtel, Fribourg, Valais, Vaud, Jura, Bern, and Geneva. The geographic distribution of these seven cantons is shown in Figure 2.



Figure 2. Geographic distributions of cantons in Switzerland (map adapted from: <https://holidaystoswitzerland.com/regions-and-cantons-of-switzerland/> (accessed on 11 November 2022).

As shown in Figure 2, the seven cantons are located in the western part of Switzerland. Based on their sociolinguistic situations, they can be further divided into two groups: monolingual cantons and bilingual cantons. Jura, Neuchâtel, Vaud, and Geneva are monolingual Francophone cantons, while Bern, Fribourg, and Valais are bilingual cantons where French and German are spoken.

3.2. Speakers

Since OFROM contains both speakers of L1 and L2 French, only native speakers of French were included in our analysis. Therefore, the speakers in the current study are all native speakers of Swiss French living in Switzerland at the time of the interview within

the seven cantons mentioned above. All the speakers were born between 1932 and 2001. As age is a factor in the current study, speakers whose year of birth was missing from their profile were excluded from the final analysis. Thus, 306 speakers were included in our final analysis. The distribution of speakers in each canton is presented in Figure 3.

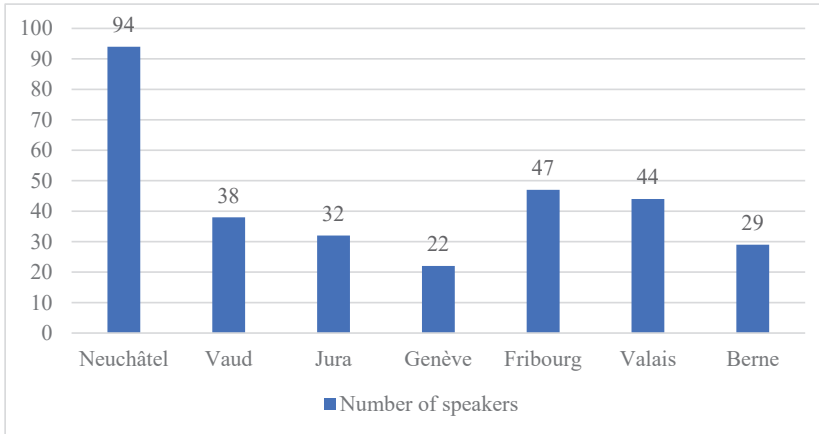


Figure 3. Distribution of speakers in seven cantons.

3.3. Tokens

In total, 587 occurrences of *genre* were identified in the OFROM corpus. Each token was first coded for its discursive status by the principal investigator. More precisely, every token is coded either as DM *genre* or non-DM *genre*. As shown in Example 1 and Example 2, the *genre* used as a noun and that can be substituted by *sorte* or *type* was coded as a non-DM *genre*, while *genre* in all other uses was coded as the DM *genre*. As shown in Figure 4, 75% of uses of *genre* were as a DM, and 25% were as a non-DM. According to our data, *genre* is mainly used as a DM in Swiss French, which suggests the grammaticalization of the particle.

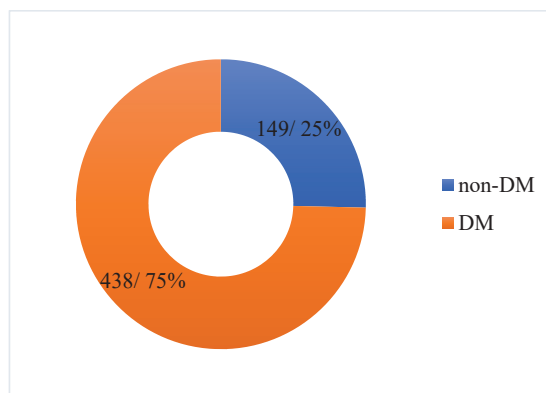


Figure 4. Distribution of the DM genre and non-DM genre in OFROM.

After being coded for their discursive status, all the tokens of *genre* were coded for their vowel quality. By listening to all the tokens and inspecting the spectrogram, we mainly coded for two realizations: *genre* without phonological reduction, realized as [ʒ^{ab}] and *genre* with phonological reduction, realized as [ʒ^{œb}]. The difference is shown in Figure 5.

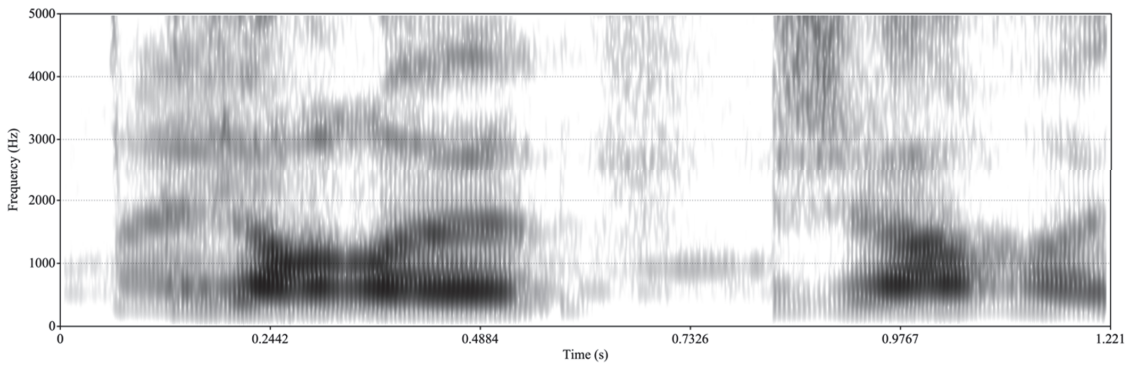


Figure 5. Spectrogram for *genre* [ʒɑ̃b] (left) without vowel reduction and *genre* [ʒœb] (right) with vowel reduction.

Figure 6 shows the distribution of *genre* realized with phonological reduction and without phonological reduction in both DM and non-DM uses. As shown in Figure 6, in its original nominal status, only 32 out of 149 uses of *genre* (21.48%) were realized with phonological reduction. However, when *genre* was used as a DM, 187 out of 438 uses of *genre* (42.69%) were realized with phonological reduction. This indicates that when used as a DM, *genre* is more susceptible to phonological erosion. This observation corresponds to the grammaticalization process of particles in general, as shown in the literature.

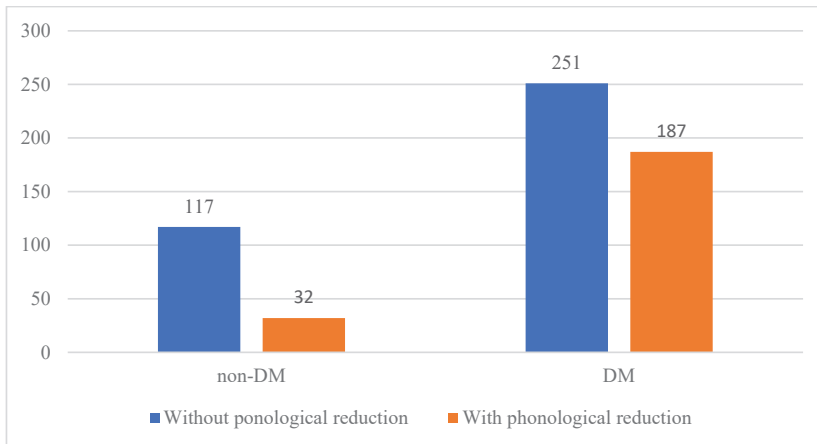


Figure 6. Distribution of phonological reduction in both DM and non-DM use.

3.4. Extralinguistic Factors

For extralinguistic factors, we looked at the age, gender, sociolinguistic situation, and social-educational status (SES) of the speakers, as shown in Table 2.

Table 2. Extralinguistic factors to be examined.

| Predictors | Levels |
|--------------------------------|--|
| Gender | female male |
| Sociolinguistic situation | monolingual non-monolingual |
| Socio-educational status (SES) | level 1 level 2 level 3 level 4 |
| Age | continuous |

For the “age” factor, we use the year of birth provided by the personal information in the corpus. For this factor, we aimed to determine, on the one hand, if the discursive use of the DM *genre* is an ongoing change in Swiss French, and on the other hand, if there is any age-grading effect on the phonological reduction when used as a DM. In either case, we should expect the “age” factor to be significant.

For the “gender” factor, we mainly want to determine whether the use of *genre* is particularly favored by any gender group. As Blondeau and Deng (forthcoming) demonstrated, in Hexagonal French, the discursive use of *genre* is favored mainly by male speakers and, therefore, is socially marked. We aimed to assess whether the same applies to the discursive use of *genre* in Swiss French.

Regarding the “sociolinguistic situation” factor, as mentioned earlier, four of the seven cantons are monolingual, and the other three are bilingual. For this factor, we mainly distinguish between two levels: monolingual and bilingual. We aimed to determine whether monolingual status facilitates the discursive use of *genre*. If the discursive use of *genre* is mainly associated with bilingual speakers, language contact might play a role. In contrast, if this use is more associated with monolingual speakers, the emergence of the discursive use of *genre* may not be the consequence of contact with another language.

As for the “socio-educational status (SES)” factor, four levels are distinguished in the corpus: Level 1: compulsory education with technical learning; Level 2: compulsory education with office learning; Level 3: high school education; Level 4: university education. Since socio-educational status is closely related to the social-economic status of the speakers, for this factor, we aimed to determine whether the discursive use of *genre* is particularly favored by speakers from any social class.

3.5. Statistical Analysis

For this article, we performed the logistic mixed-effects regression analysis in R using Rbrul (Johnson 2009; R Core Team 2021). The model distinguished the following levels for statistical significance: $p > 0.1$, not significant; $0.05 < p < 0.1$, marginally significant; $p < 0.05$, significant; $p < 0.01$, very significant; $p < 0.001$, highly significant. For the results, the model provided one p -value for each predictor (the independent variable) to indicate whether this predictor is statistically significant for predicting the dependent variable. It also provided the factor weight and log odds for each level of the predictor to indicate which levels favor/disfavor the chosen variable.

For our analysis, we first examined the dependent variable, which is *genre*, at binary classifications of the DM *genre* vs. non-DM *genre*. Second, we looked at the dependent variable, which is the DM *genre* at binary classifications of the DM *genre* with phonological reduction vs. DM *genre* without phonological reduction. For both, the fixed independent variables are extralinguistic factors presented above. All fixed factors except for age are categorical. As we used the birth year of the speakers for the age, the age factor is thus continuous. To include the mixed effects, we used speakers as a random variable. For the modeling, we performed the one-level test. Since the “participants” factor was treated as a

random variable, for the following section, we only provide the results for the fixed factors for further discussion.

4. Results

4.1. Discursive Use of genre

Overall, the logistic mixed-effects regression analysis indicated that the discursive use of *genre* is constrained by different social factors. As shown in Table 3, the gender, age, and sociolinguistic situation of the speakers are proven to be statistically significant to the discursive use of *genre* in Swiss French. In contrast, the socio-educational status of the speakers is not statistically significant to this use.

Table 3. Correlation between the discursive use of genre and extralinguistic factors.

| | | DM genre/Non-DM genre | |
|----------------------------------|----------|------------------------|-----|
| Input prob. | | 0.746 | |
| Total number | | 587 | |
| Log. likelihood | | −267.83 | |
| | F. w | % | N |
| Gender | | p = 0.000864 | |
| female | 0.668 | 82.9 | 398 |
| male | 0.332 | 57.1 | 189 |
| Age | | p = 0.00136 | |
| continuous | log odds | | |
| +1 | 0.047 | | |
| Sociolinguistic situation | | p = 0.0688 | |
| monolingual | 0.603 | 78.3 | 470 |
| non-monolingual | 0.397 | 59.8 | 117 |
| SES | | Not significant | |
| Level 1 | [0.596] | 62.5 | 8 |
| Level 2 | [0.511] | 78.9 | 76 |
| Level 3 | [0.502] | 82.2 | 129 |
| Level 4 | [0.391] | 71.4 | 374 |
| Speakers | | Random | |

As presented in Table 3, the “age” factor is statistically very significant to the discursive use of *genre* ($p < 0.01$; log odds: 0.047). As shown by the results, the larger the value of the “age” factor, the more likely a speaker will use *genre* as a DM. Since we use the year of birth of the speaker for this factor, as the value of the factor increases, the actual age of the speaker decreases. Thus, the larger the number, the younger the speaker. Therefore, the discursive use of *genre* is age-graded in apparent time with a strong association with younger speakers in the speech community. This proves that the discursive use of *genre* is indeed an ongoing change in Swiss French.

Meanwhile, the “gender” factor is proven to be highly significant to the discursive use of *genre* ($p < 0.001$; f.w.: female: 0.668; male: 0.332). Female speakers favor the discursive use of *genre*, while male speakers disfavor this use. Combined with the results for the “age” factor, our results indicate that the discursive use of *genre* is an ongoing change led by female speakers.

Regarding the “sociolinguistic situation” factor, it is only marginally significant to the discursive use of *genre* ($0.05 < p < 0.1$; f.w.: monolingual: 0.603; bilingual: 0.379). The monolinguals are more likely to use *genre* as a DM, while bilinguals are less likely to do so. This suggests that the emergence of the discursive use of *genre* in Swiss French is likely not due to language contact with other languages in the speech community. This, however, partially corroborates previous findings on *genre* in other French varieties that it is not a

calque of its English equivalent *be like* but rather an independent internal development that leads to the grammaticalization of the particle (see, for example, Cheshire and Secova 2018).

4.2. Phonological Reduction of the DM genre

As shown in Table 4, the only social factor that is statistically significant to the vowel reduction in the DM genre is the socio-educational status of the speakers. Gender, age, and sociolinguistic status are not significant to the vowel reduction in the DM genre. Our results indicated that speakers from level 3 SES are the group that tends to reduce the vowel in the DM genre. In total, 75.5% of the DM uses of genre produced by this group are found to be with vowel reduction. All other three groups disfavor vowel reduction. This suggests that the phonological reduction of genre, when used as a DM, is highly socially marked. By percentage, we can see that speakers with high SES are more likely to reduce the vowel in the DM genre, while speakers with lower SES are less likely to have vowel reduction in their DM genre.

Table 4. Correlation between the phonological reduction of the DM genre and social groups.

| | | With Vowel Reduction/Without Vowel Reduction | |
|----------------------------------|----------|--|-----|
| Input prob. | | 0.427 | |
| Total number | | 438 | |
| Log. likelihood | | −222.966 | |
| | F. w | % | N |
| SES | | p = 0.0426 | |
| Level 3 | 0.827 | 75.5 | 106 |
| Level 4 | 0.48 | 36.0 | 267 |
| Level 1 | 0.365 | 20.0 | 5 |
| Level 2 | 0.283 | 16.7 | 60 |
| Gender | | Not significant | |
| female | 0.525 | 45.2 | 330 |
| male | 0.475 | 35.2 | 108 |
| Age | | Not significant | |
| continuous | log odds | | |
| +1 | 0.014 | | |
| Sociolinguistic situation | | Not significant | |
| monolingual | 0.601 | 45.1 | 368 |
| non-monolingual | 0.399 | 30.0 | 70 |
| Speakers | | Random | |

5. Discussion: Emergence of the DM genre in Swiss French

As presented earlier in this article, the OFROM corpus we used in the current work was initiated in 2012, while the one Blondeau and Deng (forthcoming) used in their study of Hexagonal French was ESLO 2 was initiated in 2008. Based on the time of corpus construction, these two corpora offered comparable oral data on two French varieties in different countries. However, when it comes to the use of genre in these two corpora, we notice that the percentage of this discursive use in these two corpora differs significantly. As shown by our results, 75% of uses of genre were as a DM in Swiss French. Blondeau and Deng (forthcoming) reported that only 47.3% of uses of genre were as a DM in Hexagonal French. That is to say, the discursive development of genre in these two regions is not at the same pace or, at least, is not at the same developmental stage, even though in both studies, the discursive use of genre was reported to be an ongoing change in French in both countries. The discursive use of genre is much more advanced in its development in Swiss French.

At the same time, it is also interesting to see the difference in gender impact on the discursive use of *genre*. In Swiss French, it is women who led the change, while in Hexagonal French, at a comparable time, this discursive use had already spread to both genders. It is intriguing that in Blondeau and Deng (forthcoming), it is reported that in the earlier corpus ESLO 1, the use of *genre* as a DM is more associated with male speakers. Compared with our results here on gender effect, it is curious to see how the discursive use of the same particle, though an ongoing change in both countries, could be a change led by different gender groups. The original status of the DM *genre* seems to be very different in France and Switzerland. The DM *genre* in Hexagonal French is more strongly correlated with a certain gender than in Swiss French at the initial stage, while this marked status gradually disappeared over 40 years. In contrast, in Swiss French, women led the change to its advanced stage. It suggests that the DM *genre* in Swiss French and Hexagonal French are two independent processes at different stages of grammaticalization.

As demonstrated in the literature, when a particle starts to undergo the process of grammaticalization, it also begins to lose its semantic complexity. Thus, semantic bleaching is often attested at the advanced stage of development. This semantic bleaching then leads to phonological reduction since it necessarily increases the predictability of the lexical item and, in consequence, reduces the necessity of its phonological saliency. Based on our results and previous literature, it is reasonable to expect more phonological reduction in Swiss French than in Hexagonal French since the grammaticalization of *genre* is much more advanced in Swiss French. This also justifies the examination of the phonological reduction of *genre* when used as a DM in this study.

As shown in Figure 6, 42.69% of the DM uses of *genre* are realized with phonological reduction. This confirms our hypothesis that phonological reduction should be expected at the advanced stage. What is even more intriguing about phonological reduction is that in our results, it is noticed that only speakers from level 3 SES favor this variant, while speakers from other SES groups do not favor this variant and realize the DM *genre* in its unreduced form. We cannot help but ask what is special about this speaker group. Why do they behave differently from speakers from other SES groups?

Coming back to this grouping, we notice that level 3 SES speakers are mainly speakers who have completed high school education. This particularity of their language use seems more related to the theory of the “critical age”. As proposed by Labov (2001), the peak of the “critical age” is at 17 years old, when high schoolers intentionally calibrate their use of language according to their peers, a point also reckoned with by Eckert (1988, 1997). The speakers from level 3 SES are the ones that are mostly aware of the speech of their peers, while speakers from level 1 and level 2 SES may not have the chance to interact actively with their peers during the period of peer awareness since they never got that degree before they started working. As for level 4 SES speakers, as they continue to receive a university education and become more likely to use standard forms, they may intentionally use more standard linguistic forms in their speech. This result is particularly interesting because it were not attested in Hexagonal French in the literature. It supplies further evidence that the grammaticalization of the DM *genre* might be an independent process from the one in Hexagonal French, though the same use could be attested in French around the same period. Given the different developmental stages and phonological reduction degree, we have reason to believe that the two grammaticalization processes may not have a mutual influence.

Another point that needs further discussion in our results is the influence of the sociolinguistic status of the speakers on the discursive use of *genre* in Swiss French. As mentioned earlier in the article, due to the complexity of linguistic situations in Switzerland, the grammaticalization of a particle could, by all means, be the consequence of the language contact between different languages spoken in the region(s). Even though the overall percentage of *genre* used as a DM is high in the speech community, it is higher in the monolingual regions, where French is the only official language, than in the bilingual regions, where French and German are both the official languages. This gives us

reason to believe that the grammaticalization of *genre* in Swiss French is not a consequence of language contact in the Francophone regions since if it is language-contact-induced, we should see a higher percentage of discursive use in the bilingual region than in the monolingual region.

6. Conclusions

By conducting an apparent-time analysis of corpus collected in Francophone Switzerland in 2012, this article examines the variable use of the French discourse marker *genre* in the speech of 306 native speakers of Swiss French. The objective of the current study was to find further evidence of the grammaticalization of *genre* in Swiss French and discuss whether it is an independent process of grammaticalization from that in Hexagonal French documented in the literature.

The logistic mixed-effects regression analysis results revealed that the use of *genre* as a DM is indeed an ongoing change led by female speakers in Swiss French. Our analysis also proved that the phonological reduction of this particle is mainly associated with speakers with high school education. We posited that this might be related to the critical period when speakers are more sensitive to the speech of their peers. It is not the result of language contact in that speakers in French monolingual regions are more likely to use it as a DM than speakers in bilingual Francophone regions in Switzerland. Our results also demonstrated that it is at a more advanced stage of grammaticalization compared to the grammaticalization of this particle in Hexagonal French. We suggested that the grammaticalization of *genre* in Swiss French is independent of that in Hexagonal French and that it is not due to mutual influence. Our results also indicated that the DM *genre* in Swiss French has a higher percentage of phonological reduction than in Hexagonal French. This also confirms our earlier claim that the grammaticalization of *genre* in Swiss French is more advanced than in Hexagonal French.

Overall, our results supplied further evidence of the grammaticalization of the French discourse marker *genre* in Swiss French. The contribution of the current study is two-fold. On the one hand, it supplies comparable results on the grammaticalization of the same particle in two different French varieties and thus provides further evidence for the independence of the grammaticalization of a particle in different varieties of a language; on the other hand, our data also shed new light on the language variation and change of discourse markers in French and thus provide new insights into the development of the same particle in different regions being conditioned by different social factors. While the ongoing change of *genre* is led by female speakers in Swiss French, it is an ongoing change first led by male speakers, then quickly spread to both gender groups in Hexagonal French. The age-grading effect on this particle is observable in Swiss French, while it is absent from Hexagonal French.

However, some questions remain untouched by the current study. For future studies, several venues could be taken. First, as seen earlier in the literature, the DM *genre* has different discursive functions in French native speech. Very often, the discursive functions of a particle do not always stay the same in its discursive development. Some functions might emerge at a certain point in history and enter the competition with other existing discursive functions of the particle, while other functions might become obsolete and gradually disappear. In a future study, it would also be relevant to examine whether the discursive functions of *genre* change over 40 years.

Second, since we only have a general idea about the origin of DM *genre*, it is preferable to be verified in a more detailed diachronic corpus study. That is to say, it would be ideal to look into historical data to track the development of this particle diachronically. This would also provide a further comparison to its counterpart in Hexagonal French. If it is already more advanced than its counterpart in Hexagonal French, it is also possible that this discursive use began earlier in Swiss French. Looking at the historical data would also help us to understand at the initial stage how this particle enters the process of grammaticalization.

Funding: This research received no external funding.

Institutional Review Board Statement: The study did not require ethical approval. Ethical review and approval were waived for this study as it involved no experimenting with humans. The presented analyses were carried out on an already existing corpus of recordings at the Université de Neuchâtel that is publicly accessible online. The practice included informed consent from the recorded participants and reassurance by the institution that the recordings were anonymous and strictly confidential, meant to be used only for research purposes by the institution.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

- Avanzi, Mathieu, Béguelin Marie-José, Corminboeuf Gilles, Diémoz Federica, and Johnsen Laure Anne. 2012–2020. Corpus OFROM—Corpus oral de français de Suisse romande. Université de Neuchâtel. Available online: www.unine.ch/ofrom (accessed on 11 November 2022).
- Blondeau, Hélène, and Delin Deng. forthcoming. Discourse markers and social variation. In *Manuals of Discourse Markers in Romance Languages*. Edited by Maj-Britt Mosegaard Hansen and Jacqueline Visconti. Berlin: De Gruyter.
- Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.54. Available online: <http://www.praat.org/> (accessed on 9 October 2021).
- Cheshire, Jenny, and Maria Secova. 2018. The origins of new quotative expressions: The case of Paris French. *Journal of French Language Studies* 28: 209–34. [CrossRef]
- Denison, David. 2002. History of the *Sort of Construction Family*. Paper presented at Second International Conference on Construction Grammar, Helsinki, Finland, September 6–8; vol. 7, pp. 279–304. Available online: http://lings.ln.man.ac.uk/staff/dd/papers/sortof_iccg2.pdf (accessed on 11 November 2022).
- Eckert, Penelope. 1988. Adolescent social structure and the spread of linguistic change. *Language in Society* 17: 183–207. [CrossRef]
- Eckert, Penelope. 1997. Age as a Sociolinguistic Variable. In *The Handbook of Sociolinguistics*. Edited by Florian Coulmas. Oxford: Blackwell.
- Fleischman, Suzanne, and Marina Yaguello. 2004. DMs across languages. In *Discourse across Languages and Cultures*. Amsterdam: John Benjamins Publishing Company, pp. 129–47. [CrossRef]
- Fraser, Bruce. 1988. Types of English DMs. *Acta Linguistica Hungarica* 38: 19–33. Available online: <https://www.jstor.org/stable/44362602> (accessed on 11 November 2022).
- Hansen, Mosegaard M.B. 1998. The semantic status of DMs. *Lingua* 104: 235–60. [CrossRef]
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37: 1043–68. [CrossRef]
- Heine, Bernd, and Mechthild Reh. 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Buske Helmet Verlag GmbH.
- Heine, Bernd, Ulrike Claudi, and Friederike Hünemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press. Available online: <https://www.degruyter.com/database/COGBIB/entry/cogbib.5363/html> (accessed on 11 November 2022).
- Johnson, Daniel E. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3: 359–83. [CrossRef]
- Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories. *Diogenes* 13: 55–71. [CrossRef]
- Labov, William. 2001. *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford: Blackwell.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization*, 2nd rev. ed. Munich: LINCOM EUROPA. First published 1982.
- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*. Berlin: Language Science Press, p. 214. [CrossRef]
- Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scienza (Rivista di Scienza)* 6: 384.
- Mihatsch, Wiltrud. 2010. The diachrony of rounders and adaptors: Approximation and unidirectional change. In *New Approaches to Hedging*. Bingley: Emerald Group Publishing, pp. 93–122. [CrossRef]
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 11 November 2022).
- Saxena, Anju. 1995. Unidirectional grammaticalization: Diachronic and cross-linguistic evidence. *STUF—Language Typology and Universals* 48: 350–72. [CrossRef]
- Secova, Maria. 2011. Discourse-Pragmatic Features of Spoken French: Analysis and Pedagogical Implications. Doctoral dissertation, Queen Mary University of London, London, UK. Available online: <https://qmro.qmul.ac.uk/xmlui/handle/123456789/681> (accessed on 11 November 2022).

- Traugott, Closs E. 1995. The role of the development of DMs in a theory of grammaticalization. Paper presented at ICHL XII, Manchester, UK, August 13–18.
- Vincent, Nigel. 2001. LFG as a model of syntactic change. In *Time over Matter: Diachronic Perspectives on Morphosyntax*. Stanford: CSLI Publications, pp. 1–42.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Paths of Constructionalization in Peninsular Spanish: The Development of “Pues Eso”. A 20th Century Case

Shima Salameh Jiménez

Department of Spanish Language and Theory of the Literature, Faculty of Media and Communication Science, Complutense University of Madrid, 28040 Madrid, Spain; ssalameh@ucm.es

Abstract: This paper explores the evolution of the conversational formula *pues eso* in Peninsular Spanish through the framework of constructionalization, so as to describe how form–meaning pairings have been consolidated. Additionally, the Val.Es.Co. model for discourse segmentation is introduced as part of the form pole in the construction. The findings suggest that PE has become a consolidated parenthetical, procedural device during the 20th century, but that previous centuries are also key in understanding how the new functions were developed from *pues*.

Keywords: constructionalization; conversational formula; pues eso; Peninsular Spanish; 20th century

1. Introduction

The Peninsular Spanish conversational formula *pues eso* (henceforth, PE) covers five different functions: formulation (1), self-affirmation (2), back-to-topic (3), agreement (4), and closing mark (5) (Salameh Jiménez 2020b). Observe examples (1)–(5) (see Section 3 for a more detailed explanation of each function of PE):

- (1) Éste se ha ha habido que armarle fuera para poner los hilitos que suben, que bajan, que tal y luego, PUES ESO, se mete el palito, se le pone empinado, gotita de cola y, y otro y otro y otro.

You gotta assemble it outside to put the little threads that go up, down, and so on. Then, PUES ESO, you stick the little stick in, stand it up, a drop of glue, and another, and another, and another.

- (Corpus CREA, oral, Feria de Artesanía, conversación con participante, 27 March 1991. Accessed on: 28 September 2023)

- (2) B: Pero la tienes que hacer A: Sí. Ya, PUES ESO, por a veces, ¿por qué tienes que hacer las cosas? Por cubrir la papeleta. B: But you gotta do it. A: Yeah, exactly. Well, PUES ESO, sometimes, why do you have to do things? Just to cover your bases.

- (Corpus CREA, oral, Vehículo particular, conversación entre amigos, 6 July 1991. Accessed on: 28 September 2023)

- (3) C: [pues eso resulta que hacía cada ani]malada conduciendo↑ es que cuando él ya-ha venido conmigo ya había hecho las prácticas↑ y yo no lo he visto haciendo esas burradas↓.

A: es que hay algunos ((computes)) quee-quee ((()) a ni]vel habi-[de habilida]des son muy [((negados))]

C: [PUES ESO it turns out he did all sorts of crazy things while driving↑ because when he came with me, he had already done the practice↑, and I didn't see him doing those crazy things↓.

A: Some people are just ((incompetent)) when it comes to skills at that level, you know, they're really [((())]

Citation: Salameh Jiménez, Shima. 2023. Paths of Constructionalization in Peninsular Spanish: The Development of “Pues Eso”. A 20th Century Case. *Languages* 8: 289. <https://doi.org/10.3390/languages8040289>

Academic Editors: Renata Enghels, Marlies Jansengers and Tom Bossuyt

Received: 30 September 2023

Revised: 2 December 2023

Accepted: 8 December 2023

Published: 18 December 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

(Corpus Val.Es.Co 3.0. Conversation 1994 PF.1 <http://www.valesco.es>. Accessed on: 28 September 2023)

- (4) C: [ya] pero imagínate que te tiene que mirar
 B: no no no entonces no [le digo]
 A: [yaa]
 C: [PUES ESO (RISAS)]
 C: [yeah] but imagine if she has to look at you.
 B: No, no, no, then I won't [tell her].
 A: [yep].
 C: [PUES ESO (laughs)].

(Pons Bordería, Salvador (dir.): Corpus Val.Es.Co 3.0. <http://www.valesco.es>. Accessed on: 28 September 2023) (Pons Bordería 2023)

- (5) Odio las etiquetas, pero como a todo se le pone una. . . , PUES ESO. I hate labels, but since everything gets one. . . , PUES ESO.

(CREA 1995 SIERRA I FABRA, JORDI, El regreso de Johnny Pickup [España] [Madrid, Espasa Calpe 1995] Novela. Accessed on: 28 September 2023)

These five functions, in turn, are associated with various prosodic, structural, and contextual features that work together, reflecting a completed process of constructionalization (Traugott and Trousdale 2013) which remains unexplored. This paper aims to trace the diachronic evolution of PE as a construction by analyzing linguistic patterns related to pre- and post-constructional processes, including syntactic (e.g., structural integration, variation in the placement of PE), semantic (e.g., shifts in meaning, bleaching, combination with new words), and pragmatic features (e.g., acquisition of new invited inferences in new contexts). In addition, our proposal incorporates units and positions from the Val.Es.Co. model of discourse segmentation (henceforth, VAM).

The analysis seeks to achieve several research objectives, such as (i) detecting how *pues* and *eso* were combined in different contexts, (ii) under which conditions their fixation was possible, and (iii) determining whether PE results from a recent language change consolidated in the 20th century. Data were retrieved from the CDH corpus (*Corpus del Diccionario histórico de la lengua Española*, Real Academia Española) (CDH n.d.), with particular attention to written texts reproducing orality, in accordance with the concept of “imitation of orality” (López Serena 2023).

The results suggest a shift in PE towards a free, non-integrated sentence structure that preserves anaphoric–cataphoric values in novel interactive contexts. Agreement marks and self-reinforcement functions have been documented in texts published in the late 19th and 20th centuries by authors such as Pérez-Galdós, Aub, Sánchez Ferlosio, etc. Formulation and back-to-topic have also been consolidated during the 20th century, with some closing mark examples found as late as the 1990s. However, examining examples from previous centuries, specifically the 16th and 17th centuries, is crucial for observing the development of these functions through changes in morphological, structural, and pragmatic features. The VAM contributes to a better delineation of such functional development, allowing for the systematic categorization of the construction by considering combinations of functions, units, and positions. This approach aligns with previous works (Estellés Arguedas and Pons Bordería 2014; Salameh Jiménez 2020a).

From a broader perspective, the results encourage further research on the constructionalization of similar interactive formulas in Peninsular Spanish, such as *pues nada*, *pues bueno*, or *pues bien* (Salameh Jiménez Forthcoming(a)), as well as other *general extenders*, such as *y nada*, *y eso*, *no sé qué*, etc. (López Serena 2018; Llopis-Cardona 2020; see also Borreguero Zuloaga 2023, for the diachronic study of *general extenders* in Spanish). The existence of these constructions sharing similar properties in other Romance languages such as French, Portuguese, or Italian also suggests potential relationships at a higher, abstract level of a constructional network, which should be explored diachronically (Gras 2011).

The contents of this paper are distributed across the following sections: Section 2 draws upon the frameworks of grammaticalization and construction grammar (Section 2.1) and summarizes the basic units of the VAM, illustrating how they can be integrated into a constructional approach (Section 2.2). Section 3 briefly describes how PE functions in Peninsular Spanish. Section 4 is subdivided into two subsections: Section 4.1 details the methodology, addressing how the data have been handled and specifying the main linguistic parameters employed in the analysis; Section 4.2 presents the results through three sections dealing with the different stages of evolution detected in PE. Finally, Section 5 discusses the theoretical background regarding the results obtained, highlighting areas for future research.

2. Theoretical Backgrounds

2.1. Beyond Grammaticalization: Construction Grammar and Constructionalization Processes

Research in grammaticalization has been the focus of several works across languages during the 20th century (see some foundational studies, such as Givón 1971; Hopper 1991; Sweetser 1988; Traugott 1982, 1995; Traugott and Heine 1991, among others). These works describe how linguistic items develop grammatical functions in specific contexts (Hopper and Traugott 2003, p. 2), with special interest in the cognitive and functional mechanisms under the changes carried out throughout different time spans (Traugott and Dasher 2002, p. 6). Some basic features commonly assumed in grammaticalization studies are: (i) changes are unidirectional, which means that they are not usually reversible (Haspelmath 1999; Hopper and Traugott 2003, p. 99); (ii) there are generalization processes involving modifications in class words and uses in new contexts (Hopper 1991, p. 22; Arguedas 2011, p. 30); and (iii) meaning bleaching tends to be produced, which can differ depending on the linguistic item addressed (e.g., the development of verbal items, Garachana and Sansiseña 2023; the constitution of prepositional systems, Fagard and Mardale 2015; or the rise of discourse markers, Salameh Jiménez 2020b; Deng 2023; *general extenders*, Brinton 2023).

More recently, the field has been opened towards construction grammar (CxG): it describes language as a system based on form–meaning pairings (Fillmore 1988, p. 36), incorporating prosodic, morphosyntactic, and semantic features underlying their basis (Boas 2010, p. 2). Constructions are interconnected within conventional networks, exhibiting varying degrees of size and complexity along a continuum. This continuum encompasses abstract constructions governed by general rules to more specific constructions with variable morphosyntactic and semantic properties (Fried and Östman 2004, pp. 18–22). Consequently, grammar (or language, from a broader perspective) is viewed as a holistic framework wherein no single level of grammar is autonomous or central. Instead, semantics, morphosyntax, phonology, and pragmatics collaborate within a construction (Traugott and Trousdale 2013, p. 3). Some recent studies have applied CxG in interactive, conversational contexts so as to address different grammatical and discursive elements (e.g., discourse markers, conversational formulas, etc.). These devices can be regarded as procedural constructions based on the combination of formal, functional, and contextual properties. This branch of CG is known as interactive construction grammar (ICG) (Gras 2011; Fischer and Alm 2013).

The diachronic approach to CxG is the so-called constructionalization: this framework explains how new constructions are created and gradually developed as a result of language change (Trousdale 2014; Petré 2020). Constructionalization posits that language change occurs when “new patterns come to be entrenched not only in individual minds (‘innovations’) but come to be shared and entrenched within a community of speakers (‘changes’)” (Traugott 2020, p. 129). Specifically, constructionalization arises from conventionalization, signifying that mere innovations are insufficient to initiate the constructionalization process. Conventionalized constructions involve the emergence of new type nodes with novel syntax or morphology and coded meaning within the linguistic network of a speech community, as well as “changes in schematicity, productivity, and compositionality” (Traugott and Trousdale 2013, p. 22).

Constructionalization involves two general stages for change which can occur before or after the process of constructionalization itself: pre-constructionalization and post-constructionalization (Traugott 2015). Pre-constructionalization refers to small and local changes in specific contexts which can lead to the creation of a new micro-construction through gradualness (i.e., through small-step neoanalyses triggering changes; Traugott and Trousdale 2013, p. 22). Pre-constructionalization processes are usually associated with small semantic distributional changes (do Rosário and de Oliveira 2016, p. 244), without expecting big formal or syntactic modifications in the form–meaning pairing (Enghels and Garachana 2021). For their part, post-constructionalization processes reveal a stage in which a new meaning is completely associated with a new form, thus emerging a new consolidated construction (e.g., the periphrasis *ir a + inf* in Peninsular Spanish; Enghels and Garachana 2021, p. 325). Some main post-constructional features are the increase in frequencies of use, the employment of the construction with new words and contexts not previously allowed, and the change by which the new construction can become the basis for further developments in other, new, or derived constructions. Also, constructions at this stage typically involve expansion of collocations, morphological and phonological reduction (e.g., *a lot of large quant* experimenting expansion and phonological reductions after first pre-constructionalization processes; Traugott and Trousdale 2013, p. 27).

In this paper, we adopt an analysis model based on the [F] (form) and [M] (meaning) correlation from Croft’s 2001 model (Traugott and Trousdale 2013, p. 8; Zhuo and Peng 2016; see also Traugott 2022). Considering the orality of PE, although the total amount of data retrieved may seem small in specific timespans, it allows for the recreation of the change pattern developed by PE to serve as the basis for future research. This constructional analysis is combined with the Val.Es.Co. model (VAM). The next subsection provides a brief description of the VAM units employed in the analysis.

2.2. The Val.Es.Co. Model of Discourse Segmentation (VAM): A Complementary Tool

The VAM (Val.Es.Co. Group 2014) draws on various approaches, including conversation analysis and discourse analysis (see some foundational references in Briz et al. 2003). This model, designed for discourse segmentation, consists of eight units, hierarchically and recursively related: *subact, act, intervention, exchange, turn, turn-alternance, dialogue, and discourse*. These units are categorized into three orders, informative, structural, and social, and distributed across two levels, monological and dialogical, depending on whether they reflect the speaker’s individual production or the way speakers interact (see further details in Pons Bordería 2022).

Additionally, the VAM units occupy four positions: *initial, medial, final, and independent*. These positions are relative, implying that a specific position can be associated with different units depending on the linguistic phenomenon analyzed: for example, discourse markers can occupy the same position (e.g., *initial*) in different units (e.g., *act, intervention, dialogue*) depending on the function addressed (e.g., *turn-taking, reformulation, etc.*). This relative nature of positions allows for a high level of accuracy in the analysis, as well as for a closed grid crossing units and positions in a visual way (see Table 1):

Table 1. Units and positions in the Val.Es.Co. model.

| | | Units | | | | | | | | | | |
|-----------|-------------|--------|-----|-----|-----|-----|-----|------------------|----------------|----------|----------|-----------|
| | | Subact | | | | | Act | Intervention | | Exchange | Dialogue | Discourse |
| | | DSS | SSS | IAS | MAS | TAS | | INITIATIVE (I-I) | REACTIVE (I-R) | | | |
| Positions | Initial | | | | | | | | | | | |
| | Medial | | | | | | | | | | | |
| | Final | | | | | | | | | | | |
| | Independent | | | | | | | | | | | |

From a constructional perspective, the VAM enables a systematic description of discursive constructions, as units and positions can be integrated as part of the formal pole in constructions, connected, in turn, to the meaning pole as the basis of the functions they elicit (see Croft’s model in Figure 1). The meaning pole relies on the notion of *conventionalization*, which can be measured by changes in frequencies or by how a node in a consolidated construction leads to new constructions and productive subschemas (Flach 2020, p. 18).

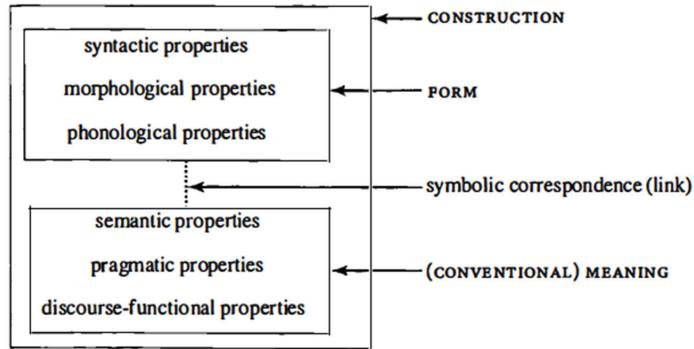


Figure 1. Constructional analysis proposed by Croft (Croft 2001, p. 18).

That said, transformations in units and positions in the constructional description can be seen as changes in the formal pole involving changes in the meaning pole. For the purposes of this paper, we will focus specifically on dialogues, interventions, acts, and subacts applied to examples showing a big degree of orality.¹

2.2.1. Intervention

Intervention is the maximal, structural unit resulting from the content uttered by speakers in a conversation. Interventions are associated with turns, which are produced once participants accept and validate the speaker’s intervention (Pons Bordería 2022, p. 79). They can be classified into three types: *initiative* (iI), *reactive* (rI), and *reactive–initiative* (r-iI). Initiative interventions trigger responses from other speakers, which are reactive interventions. When a reactive intervention triggers, in turn, new responses, it becomes reactive–initiative (Espinosa 2016, p. 15). Example 6 presents these three types:

- (6) iI P: he ido a la tienda esa/bueno↓/aa-/a cobrarte la lotería//y ee-lo que me han dao/de la lotería
 - r-iI T: ¿esto?
 - rI P: (claro)/o-ochocientas pesetas y doscientas que llevaba yo
 - iI P: I went to that store/well↓/uu-/to collect your lottery winnings//and uu-what they gave me/from the lottery.
 - r-iI T: This?
 - rI P: Yeah (of course), e-eight hundred cents and two hundred that I had.
- (Val.Es.Co. Group 2002. Accessed on: 28 September 2023)

Interventions are composed by smaller and larger units. Specifically, different interventions combine to form exchanges, which then give rise to dialogues.

2.2.2. Dialogue

Dialogue is a high-level structural unit which results from the combination of different interventions and exchanges made by different speakers. Dialogues are based on the notion of topic change, also related to interventions in the VAM: a new topic involves an initiative intervention at the beginning and a reactive intervention at the end. This can be represented schematically as follows (see Figure 2):

iI- -riI--riI--riI--riI- ... --riI -rI

Figure 2. Succession of initiative and reactive interventions within a specific dialogue (Pons Bordería and Fischer 2021, p. 106).

Dialogues also require interventions to be recognized as turns (i.e., they are based in turn-alternance). All these features of dialogue can be observed in examples 7 and 8, in which a dialogue starts with a first initiative intervention (see 7), and a new dialogue is introduced after closing the previous one with a last reactive intervention (see 8):

- (7) **A: me encanta la gente que sale a correr// o va al gimnasio que tienen pantalón corto yo cuando salgo a correr voy con pantalón largo no puedo**
 B: ah pues yo con-yo soy de esa gente yo cuando voy al gimnasio como está al lado de mi casa voy ya con los pantalones cortos
A: I love people who go for a run or to the gym wearing shorts. When I go for a run, I wear long pants; I just can't do shorts.
 B: Oh, well, I'm one of those people. When I go to the gym, since it's right next to my house, I just go in shorts.

(Corpus Val.Es.Co 3.0. Conversation 2017 PT.24 <http://www.valesco.es>. Accessed on: 28 September 2023)

- (8) **A: o en la caseta de la piscina ¿¿no?? (RISAS)/ya ves nano**
 B: aay y tú que-y no sé↑ ¿¿qué parida dijiste tú dee→del día que estábamos ahí [tiraos en el césped?]?//no ((sé que)) dijimos-dijiste sé qué del
 A: ¡[jah!! sí me acuerdo]
 A: yaa pero no me acuerdo lo que dije ya
B: por cierto laa laa Miriam esa nano ¿¿se llama Miriam la hermana de Verónica?? sí
 A: sí↓ yo creo que sí
 A: Or at the pool booth, right? (laughs) You know what I mean, bro.
 B: Oh, and you, what-I don't know, what nonsense did you say about-about that squash thing from abroad that day when we were lying on the grass? We said-you said something about squash.
 A: Oh yeah, I remember now. Well, I remember the moment, but I don't remember exactly what I said.
B: By the way, that Miriam, the one who's Veronica's sister, right?
 A: Yeah, I think so.

(Corpus Val.Es.Co 3.0. Conversation 1994 PF.8 <http://www.valesco.es>. Accessed on: 28 September 2023)

Interventions, as the basis of exchanges and dialogues, depend on their immediate, monological constituent in the VAM: the act.

2.2.3. Act

The *act* is the monological, structural unit, hierarchically dependent of interventions (Briz et al. 2003, p. 31; Pons Bordería 2016, p. 547). Acts in the VAM are defined by the combination of three main features: illocutionary force, propositional content, and intonation groups. The illocutionary force is recognized as a primary, main identifiable feature which can be completed with semantics and prosody as secondary features (Pons Bordería 2022, p. 95). Example 9 represents how acts are segmented in the VAM:

- (9) B1: #pues yo/es que/no sé/yo es que/yo sí que estoy segura#
 A1: # mm ¿estás segura?/ ¿seguro? /o sea ¿lo tienes claro?#
 B2: #sí#
 A2: #(pues yo no)#
 B1: #Well, you know, ehh-I don't know, I mean, I really am sure. #
 A1: #Are you really sure? /I mean, like, really sure? #

B2: #Yes#
 A2: #Well, I'm not#.

(Example retrieved from the DPDE, description of *o sea*)

Example (9) presents different interventions composed by acts: B1 produces a statement; A1 asks through an interrogation; B2 produces an affirmation; and, last, A2 also makes another statement. These different illocutionary forces are the core of each intervention produced by Speakers A and B. All these acts are annotated by using the symbol “#”.

Finally, acts are made up of lower units, the subacts. The relationship between acts and subacts establishes a bridge between the structural and the informative levels in the VAM.

2.2.4. Subact

Subacts are semantic–informative units which constitute the minimal informative, monologic level in the VAM (Val.Es.Co. Group 2014, p. 54). These units are identified by prosodic and semantic marks. Subacts are key units in the VAM since they represent the space where discourse markers, constructions, etc., can be analyzed (Pons Bordería 2022, p. 101). Depending on the type of semantic information they introduce (i.e., causes, conditions, consequences, etc.), subacts can be substantive (SS) or adjacent (AS), as observed in the following examples:

(10) # {DSS esta es de cuarenta y dos→esta también tiene sus años DSS}# # {DSS iba yo al cole→//y le dimos→[una sor]presa a mi padre→porque yo me acuerdo↑///(4”) que la compró mi madre↑/ se la compró mi madre a mi tío Christian↑ # {DSS entonces [la puso aquí] DSS}# # {SSS yy y era que mi padre iba a por mí aa inglés^ SSS} (. . .) # {DSS This one’s forty-two years old, too. DSS}# # {DSS It’s been around for a while, DSS}# # {DSS I was going to school, and we surprised my dad, because I remember that my mom bought it.///(4”) My mom bought it from my uncle Christian, DSS}# # {So DSS She placed it here, DSS}# # {SSS And, and it was because my dad was picking me up from English class SSS} (. . .)

(Corpus Val.Es.Co 3.0. Conversation 2018 PT.26.S10. <http://www.valesco.es>. Accessed on 28 September 2023)

Example (10) presents different SS based on substantive content, which can be more relevant or secondary within the intervention produced. Depending on such relevance, SS can be classified as directive substantive subacts (DSSs) and subordinate substantive subacts (SSSs). DSSs are highly informative; SSSs are semantically and informatively dependent on DSSs. In this case, the SSS expands the information included in the previous DSS.

SS are combined with AS. Adjacent subacts are namely based on procedural meaning and can also be subclassified into further categories: textual adjacent subacts (TASs), modal adjacent subacts (MASs), and interpersonal adjacent subacts (IASs). The meaning behind AS does not affect the propositional content underlying acts (i.e., their logical form; Pons Bordería 2016). As a result, they could be deleted from any example without provoking a loss of conceptual meaning as would happen with SS. The following example includes different types of AS:

(11) A: # {MAS hombre MAS} {TAS pues TAS} {DSS no pero {TAS es que TAS} a Josmi se lo piden seguro, {IAS ¿no? IAS} DSS}#
 A: # { MAS Well MAS} {TAS then TAS} { DSS no→ but {TAS the thing is TAS} they probably ask Josmi too, } {IAS right↑? IAS} DSS} #

(Corpus Val.Es.Co 3.0. Conversation 2011 PT.S3 <http://www.valesco.es>. Accessed on 28 September 2023)

In Example (11), TASs organize and distribute the flow of speech; they reflect the relationship between the ideas expressed and the discourse. MASs allow speakers to modalize the discourse they are related to; they highlight the relationship between the

speaker and his/her own discourse. Last, IASs show how speakers interact to each other in a discourse; then, they highlight the relationship between speaker and hearer(s).

3. General Features of PE in Peninsular Spanish

Peninsular Spanish PE has been commonly described as a conversational formula derived from the Spanish discourse marker *pues* (Álvarez 1999; Briz 1993; Pons Bordería 1998; see also Iglesias 2000 for a diachronic approach until the 15th century). As argued in Salameh Jiménez (2020b, pp. 110–11), the pragmatic uses of *pues* (e.g., formulation, or cataphoric–anaphoric values introducing non-preferred answers; Briz 1993, p. 156) have triggered new, fixed combinations with other linguistic items (e.g., *pues nada*, *pues bueno*, *pues bien*, *pues vale*; Salameh Jiménez Forthcoming(a)). Also, PE is determined by the deictic, neutral pronoun *eso*, which creates a cognitive space leading to non-tangible elements such as ideas or concepts without explicit mention (De Cock 2013, p. 11). In particular, the neutral pronoun *eso* is related to vagueness processes produced in the so-called *abstract region* (Achard 2001), in which any discursive content could be placed (i.e., fillers, formulative devices, etc.).

All these features have contributed to the fixation of *pues* and *eso* even earlier than other formulas based on a schematic abstraction of *pues* (Salameh Jiménez Forthcoming(a)). In PE, some original properties from both items have been integrated into its meaning basis to some extent; however, some new properties have been acquired in new contexts (Croft and Cruse 2008, p. 350). Thus, the functions of PE cannot be addressed as only the compositional combination of the meanings behind the original forms. This formula presents five different functions (see Section 1): formulation, reinforcement, back-to-topic, agreement, and closing mark (Salameh Jiménez 2020b).

Formulations with PE are produced when the speaker discusses a specific topic and further information needs to be added to make the conversation progress. Whether the speaker cannot find immediately the most precise words, PE establishes a pause to search for the new content and also allows them to maintain the floor in the conversation, like in Example (12):

- (12) 523 A: luego tengo un grupo de amigas también que con ellas también hago→/pues eso↓/salimos↑/¡VA!/NOS VAMOS A JÁVEA/porque una tiene allí un apartamento↑/vamos allí§
 523 A: Then I have a group of friends too, and with them, we also do→/PUES ESO ↓/we go out↑/WE'RE HEADED TO JÁVEA/because one of them has an apartment there↑/we're going there§

(Corpus Valesco 2.0, Conversation 1, Interventions 519–527. Accessed on: 28 September 2023)

Reinforcement allows speakers to emphasize the illocutionary force associated with their statement, which constitutes a propositionally complete content, as shown in (13). Most parts of reinforcement contexts are also related to self-conclusions, especially employed in the final position.

- (13) F: y te dan zapatillas/zapatillas también para levantarte
 P: pero si tienes por ejemplo dolor de lumbago/pues eso
 F: And they also give you slippers/slippers to wear when you get up
P: But if you, for example, have lower back pain/PUES ESO

(Corpus Valesco, printed version 2002, p. 288)

The back-to-topic use of PE is very common in colloquial conversations: speakers tend to make (micro-)changes of topic in the flow of a conversation and, later, they decide to come back again to one of the topics previously addressed by the group. In these cases, the speaker also employs PE as a turn-taking device, as observed in (14):

- (14) A: [María Jesús ¿¿te acuerdas cuando] fuimos a-?? aa-eras tú ¿¿no?? la que vi[nistes conmigo (GRITOS)] (RISAS) [(())(ENTRE RISAS)]

B: [sí] [sí sí sí]
 C: [que no se enter]aba/[es igual]
 A: (()) [(())] no lo vi pasar ni nada y bueno es que me entró el ataque [(de risa) (())]
 C: [pues eso resulta que hacía cada ani]malada conduciendo↑ es que cuando él ya-ha venido conmigo ya había hecho las prácticas↑ y yo no lo he visto haciendo esas burradas↓
 A: es que hay algunos ((computes)) quee-quee [(()) a ni]vel habi-[de habilida]des son muy [(negados)]
 A: [Maria Jesús, do you remember when] we went to-?? aa-it was you, right?? who came with me (SHOUTS) (LAUGHTER) [(()) (AMIDST LAUGHTER)]
 B: [yes] [yes yes yes]
 C: [he had no idea]/[it doesn't matter]
 A: (()) [(())] I didn't even see him pass by or anything, and well, I burst into laughter [(laughing) (())]
 C: [PUES ESO it turns out he did all sorts of crazy things while driving↑ because when he came with me, he had already done the practice↑, and I didn't see him doing those crazy things↓
 A: Some people are just ((incompetent)) when it comes to skills at that level, you know, they're really [(())]

(Corpus Val.Es.Co 3.0. Conversation 1994 PF.1 <http://www.valesco.es>. Accessed on: 28 September 2023)

In dialogical contexts, PE can work as an agreement mark, in some cases even functioning itself as a response intervention. It becomes an explicit expression of acceptance of what is said by another speaker, as in (15):

- (15) B: bueno la verdad es que en fallas en
 A: pues eso→[que nos lo hagamos mirar]
 B: Well, the truth is that the festivals are <fsr t="really">really</fsr>→
A: PUES ESO→[we should get it checked out]

(Corpus Val.Es.Co 3.0. Conversation 2011 PT.S3 <http://www.valesco.es>. Accessed on: 28 September 2023)

Last, the closing mark in PE highlights how speakers end a topic they were talking about and allows for introducing a new one, which can be made by the same speaker or by another speaker. In these contexts, PE can be combined with other discourse markers, such as *nada* or *bueno*, as shown in (16):

- (16) A: ¿¿y quinientas pelas te ha costado ese paqueti[to!]?
 B: ¡[no!][↑] que llevaba quinientas pelas[↑] he comprao tabaco[↑]/y he comprao los folios y me han costadoo doscientas y pico doscientas treinta.[↓] el paquete el paquete de folios pues→
 C: mm
 B: pues eso// y nada
 A: ¿¿tú te quedas a comer Nico??
 C: no[↓] yo me voy ya
 A: And did that pack of paper cost you five hundred pesetas?
 B: [No!][↑] I had five hundred pesetas[↑], I bought cigarettes[↑], and I bought the paper, and it cost me two hundred and something, like two hundred and thirty[↓] for the pack of paper, well→
 C: Mm.
 B: PUES ESO//y nada
 A: Are you staying to eat, Nico??
 C: No.[↓] I'm leaving now.

(Corpus Val.Es.Co 3.0. Conversation 1996 PF.11 <http://www.valesco.es>. Accessed on: 28 September 2023)

As said, these five functions have been developed by combining form and meaning changes throughout time, abandoning compositional processes. To detect how these changes happened, we adopt a constructionalization approach incorporating some useful descriptive notions, such as the pre- and post-constructional changes (also, see Brinton's 2023 recent work on the evolution of *what-general extenders* in English).

4. Results

4.1. Data Compilation and Analysis Parameters

The diachronic study of orality through written texts is commonly assumed by researchers, given that recreations of interaction can be found in literary texts, which allows for understanding how real variation was in the past (Del Rey Quesada 2020). Thus, a play presenting an oral use of a linguistic item could reveal that this specific linguistic item was being diffused or conventionalized in a time period (see also López Serena and Rivera 2018).

All the samples of PE in Peninsular Spanish have been compiled from the *Corpus del diccionario histórico de la lengua Española* (CDH) from the Real Academia Española (RAE), which is composed of 355,740,238 items distributed as follows: (a) the nuclear CDH section, covering more than 32 million texts from Peninsular Spanish; (b) the CORDE section, which includes texts from the *Corpus Diacrónico del Español* since the 12th century to 1975; and (c) the CREA section including texts from 1975 to 2000. The fact that the CDH includes a big number of materials from CORDE and CREA supports its representativity for the purposes of our study.

All the examples have been manually extracted and classified depending on the degree of fixation in PE (i.e., cases without fixation (17) and cases with fixation (18)):

(17) —¡Deshonra! —exclamó el padre con asombro—, deshonra ha dicho usted, señora. **Pues eso** sí que es cosa grave.
—“Disgrace!”—exclaimed the father with astonishment—, “disgrace you have spoken, madam. For that is indeed a serious matter.”

(1871 PÉREZ GALDÓS, BENITO, El audaz. Historia de un radical de antaño [España] [CORDE, Madrid, Real Academia Española 2003] Novela)

(18) **Pues eso**. Sólo queda por averiguar si a las autoridades alguien les ha explicado que en propia meta también cuentan.
Pues eso. The only thing left to find out is whether someone has explained to the authorities that own goals also count.

(25 May 1996, El Mundo, 25 May 1996 [España] [Madrid, Unidad Editorial, 25 May 1996] Arquitectura)

Examples without fixation of *pues* and *eso* have also been analyzed so as to detect first documentations of non-compositional uses of PE, especially in those cases related to bridging contexts in the language change process. This part of the study has been key also in detecting pre-constructional changes previous to the complete constructionalization. The analysis is based in the following linguistic properties: morphosyntactic (syntactic integration or liberation, scope, including positions and units, as well as changes from the monological to the dialogical level), semantic (bleaching in meaning, changes in meaning in *pues* and *eso*, etc., acquisition of new degree(s) of (inter)subjectivity), and pragmatic (new contextual invited inferences). The VAM will be considered part of the form pole within the construction.

4.2. Results

A general approach to the quantitative data suggests some relevant information from frequencies. There are a total of 733 tokens of *pues* and *eso* in the CDH, including both compositional and constructionalized cases. Table 2 summarizes the distribution of the

733 cases per century and the normalized frequencies from the CDH corpus (all the words per century):

Table 2. Total amount of compositional/constructionalized cases of *pues* and *eso* in the database, with normalized frequencies from the CDH corpus (Spain).

| | #Absolute | #Normalized (/1,000,000 Words) |
|--------------|-----------|--------------------------------|
| 15th Century | 4 | 0.16 |
| 16th Century | 40 | 0.84 |
| 17th Century | 91 | 2.97 |
| 18th Century | 16 | 1.48 |
| 19th Century | 115 | 3.33 |
| 20th Century | 467 | 4.34 |

A focus on the data reveals a first, intuitive idea of the dialogicity behind this conversational formula: a total of 170 tokens of 733 in the database are related to dialogical texts. This means that the combination of *pues* and *eso*, fixed or not, has appeared almost 200 times in the database in contexts reproducing interaction in novels and plays, specifically introducing answers to the interventions made by others. As observed in Table 3, a clear increase in PE in these dialogical contexts is detected since the 19th century, which would reveal that the constructionalization of this formula would be apparently completed in these centuries:

Table 3. Cases of *pues* and *eso* related to dialogical contexts in the database (Spain).

| | #Absolute | #Normalized (/1,000,000 Words) |
|--------------|-----------|--------------------------------|
| 15th Century | 1 | 0.04 |
| 16th Century | 2 | 0.042 |
| 17th Century | 2 | 0.065 |
| 18th Century | 1 | 0.09 |
| 19th Century | 43 | 1.24 |
| 20th Century | 121 | 1.12 |

Nevertheless, an automatic search alone is insufficient to draw definitive conclusions. The analysis of the remaining 562 tokens is essential to understand why PE became established as a construction and why its scope expanded from monological to dialogical contexts. It is also crucial to examine the earlier stages of the construction, even before it functioned as a parenthetical device, in order to trace the pattern by which a connective followed by a neutral deictic changed into a fixed response and how it acquired textual, modal, and interactive uses.

That said, the PE data reveal three primary stages in the constructionalization process: non-constructionalized PE (Section 4.2.1), constructionalization in progress (Section 4.2.2), and constructionalized PE (Section 4.2.3). These stages, in turn, are associated with pre- and post-constructionalization processes. Pre-constructional changes are initiated in the first and, notably, in the second stage, while post-constructional processes are a part of the third stage, where the constructionalization of PE appears to be completed for at least four out of five values.

4.2.1. First Stage: Non-Constructional PE

As anticipated, the initial instances of PE found in the database exhibit a compositional, non-constructional meaning. This implies that *pues* and *eso* are not part of the same integrated structure, which can be verified through various formal linguistic tests. Consider the following example from the 15th century:

- (19) e como dixo meliangans pues yo he lo mejor dela batalla **pues eso** que tu dizes dixo el Rey non te vale cosa que vien vemos todos como es entonçes

dixo meliangas yo buscare mi derecho por do quier que pudiere pues de dicho la batalla tullir non me podeis e desde aqui digo a lançarote que si al presente dexa la batalla que lo [...] sentençado entonçes tiro (...) I have the best of the battle," well, what you say, said the King, doesn't help you at all. We all see how it is. Then Meliangans said, "I will seek my right wherever I can, for you cannot stop me from leaving the battle." And from here, I tell Lancelot that if he leaves the battle now, they will consider it a sentence. Then he threw (...)

(c1414 ANÓNIMO, Traducción de Lanzarote del Lago [España] [Harvey Sharrer, Madison, Hispanic Seminary of Medieval Studies 1999])

In (19), there is a combination of *pues* and *eso* in which *pues* works as a conjunction, relating the utterance it introduces to a previous cause expressed through the preceding utterances, while *eso* functions as a neutral pronoun, serving as a deictic reference to all previously mentioned information (i.e., functioning as an anaphoric device). Morphosyntactically, *eso* is fully integrated into the sentence, as evidenced by its predication relationship with the verb "es" (to be) and other complements. It admits the substitution of the deictic "eso" by other neutral devices that refer to the preceding content (e.g., *pues lo que tú dices*, *pues aquello que tú dices*, *pues la cosa que tú dices*, etc.). Importantly, *eso* is used within lengthy sentences and even subordinate clauses. Semantically, there is no bleaching or reanalysis that leads to new pragmatic uses of *pues + eso*; instead, they retain their original, fundamental meanings that are not associated with a potential schematic construction based on *pues*. Finally, according to the VAM, *pues* and *eso* would be categorized within the directive substantive subact (DSS) with scope over the following subordinate substantive subact (SSS). This reflects their non-liberation or non-consolidation as a construction.

Example (20) also presents uses of PE which are part of the sentence level, without syntactic complete freedom:

(20) (...) aunque pongamos toda la firmeza posible en tener el pensamiento fijo en Dios cuando estamos orando y aun celebrando, al mejor tiempo se nos va de casa sin licencia, y desaparece; y tornándolo a traer, luego torna a desaparecer, y como una anguila se nos cuela por entre las manos. **Pues eso** es lo que principalmente hace dificultosísimo este ejercicio. (...) even if we put all the firmness possible into keeping our thoughts fixed on God when we are praying or even celebrating, at the best time, it escapes our house without permission and disappears. And when we bring it back, it immediately disappears again, slipping through our hands like an eel. Thus, this is what primarily makes this exercise extremely difficult.

(1554 GRANADA, FRAY LUIS DE, Libro de la oración y meditación [España] [Álvaro Huerga, Madrid, Fundación Universitaria Española-Dominicos de Andalucía 1994] Religión)

The difference regarding cases like (19) is that *pues* introduces a conclusion derived from the previous utterances (i.e., it does not only mark a cause–consequence relationship within a sentence) and that it works as a connective mark which could be deleted from the sentence without a loss of meaning:

(...) y tornándolo a traer, luego torna a desaparecer, y como una anguila se nos cuela por entre las manos. **Eso** es lo que principalmente hace dificultosísimo este ejercicio.

And when we bring it back, it immediately disappears again, slipping through our hands like an eel. This is what primarily makes this exercise extremely difficult.

Nevertheless, these examples are still compositional and do not show any degree of constructional (semi)fixation. The database retains a total of ca. 170 compositional contexts of *pues* and *eso*, which is expected since the connective *pues* has not disappeared or been replaced by other constructions that it may have triggered (indeed, *pues* and other

derived constructions, such as PE, *pues nada*, or *pues bien* would not be placed at the same level within a constructional network). Table 4 provides a summary of their distribution throughout centuries in the corpus, including normalized frequencies:

Table 4. Cases of *pues* and *eso* in the first, compositional stage in our database (Spain).

| | #Absolute | #Normalized (/1,000,000 Words) |
|--------------|-----------|--------------------------------|
| 15th Century | 2 | 0.08 |
| 16th Century | 21 | 0.44 |
| 17th Century | 42 | 1.37 |
| 18th Century | 5 | 0.46 |
| 19th Century | 25 | 0.72 |
| 20th Century | 94 | 0.87 |

4.2.2. Second Stage: Constructionalization in Progress

The second stage encompasses initial, bridging uses of *pues* and *eso*, involving some reanalysis processes that alter specific aspects of meaning and pragmatic values. Continuative and illative original values of *pues* are key in this second stage so as to lead to pre-constructionalization constructional shifts. Notably, during the 16th and 17th centuries, agreement and reinforcement marks appear to be activated in contexts where the construction is not yet fully established. Of particular relevance for the analysis is the combination of *pues* and *eso* with specific verbal constructions and, in a more fixed form, with *mismo* (adj.; *pues eso mismo*) in dialogic contexts, where it serves as a response, or in monological contexts, reinforcing the speaker’s own statements. A closer examination of these initial uses suggests a formal–functional specialization of PE, contributing to a new, less compositional meaning that precedes new parenthetical uses (see Section 4.2.3).

4.2.2.1. Agreement Mark

Examples (21) and (22) illustrate the emergence of the first continuative values marked by the agreement feature, represented schematically as PE + V:

(21) Latancio.—**Pues eso juzgó**, mávalo allí como él se lo dixo, ca dos nobles mejores que a sí mató. Latancio.—Well, that’s what he decided; kill him right there, just as he told you, because he killed two nobles better than himself.

(1471–1476 [finales del s. XV] GARCÍA DE SALAZAR, LOPE, *Istoria de las bienandanzas e fortunas [España]* [Ana María Marín Sánchez, Madrid, Corde 2000] *Historia*)

(22) —**Pues eso es así**, yo y tu madre cenaremos estos dos. Tú cena el que heciste, que quien eso sabe razón es que eso cene.
—Well, that’s right. Your mother and I will have these two for dinner. You have the one you made, because whoever did that should be the one to eat it.

(1574 SANTA CRUZ DE DUEÑAS, MELCHOR DE, *Floresta española [España]* [Maxime Chevalier, Barcelona, *Crítica* 1997] *Relatos*)

In these instances, *pues eso* remains integrated at the sentence level, albeit with a degree of syntactic liberation reflected in certain morphosyntactic features. In contrast to examples (19) and (20), *pues eso* now reduces the number of verbal complements and limits its compatibility with subordinate clauses (e.g., *pues eso que*). As a result, the length of the utterance containing *pues eso* is shorter. Both *pues* and *eso* are used in a dialogical context, where *pues* serves to introduce a continuation of the previous discourse through a response.

These dialogical contexts may trigger reanalysis derived from the continuative original value of *pues*, employed specifically to accept something stated by others or provide further information to a previous intervention (Briz 1993). However, it is essential to note that the constructionalization process is not yet complete. The neutral pronoun *eso* can function

as both a subject and a direct object, indicating a lack of morphosyntactic freedom and suggesting that its removal would result in a loss of grammatical meaning. Additionally, segmentation using the VAM units confirms that *pues* and *eso* do not yet constitute a fully adjacent subact (AS). “Pues” can be tagged as a textual adjacent subact (TAS), while the remainder of the sentence, including *eso*, belongs to the directive substantive subact (DSS). This pattern is consistent across other examples in the corpus, indicating some degree of compositionality.

In summary, the combination of form and meaning parameters reveals an ongoing change. The possibility of new pragmatic interpretations stemming from *pues* and their expression through *eso* within a broader paradigm of neutral devices signifies a change, as reflected in formal parameters (e.g., morphosyntactic reduction, verbal restrictions). Furthermore, there is a positional and scope transformation evident in the VAM units: examples from the initial stage were typically positioned within the medial part of the DSS, whereas the new uses in agreement contexts are now placed at the initial position of the intervention, creating a new context for the use of *pues*.

Similarly, (23) shows an instance of first uses of *pues* and *eso* within a more specific agreement construction:

(23) A: Que ayude el sol no lo niego, mas para engendrar un yo, otro yo es fuerza: que el fuego dará calor al que obró el ser que me forma luego.

B: **Pues eso mismo te digo:** que el sol que una vez llegó a estar, Rosaura, conmigo, en mí misma te engendró.

A: That the sun helps, I do not deny it, but to beget one “I,” another “I” is necessary; for the fire will give heat to the one who worked the being that later forms me.

B: Well, I tell you the same: the sun that once came to be, Rosaura, with me, engendered you in myself.

(c1612 VEGA CARPIO, LOPE DE, El animal de Hungría [España] [Biblioteca Virtual Miguel de Cervantes, Alicante, Universidad de Alicante 2002] Verso)

Morphosyntactically, both elements are still integrated into the sentence structure and allow for additional verbal complements. However, there is a reduction in length. Additionally, combinations with other verbs appear less frequent, resulting in clearer, explicit agreement, functioning as an echoic marker. Semantically, the meaning of *pues* and *eso* becomes more constrained due to the explicit agreement expressed through devices like *mismo*, which unambiguously refer to the preceding information in conjunction with the neutral deictic *eso*.²

From a pragmatic perspective, this structure can be interpreted in two ways due to the illocutionary force expressed: a literal reading, where Speaker B is simply echoing Speaker A, and a non-literal reading that triggers an invited inference of acceptance, further reinforced. This functional specialization in using *pues* and *eso* restricts the introduction of refusals in these contexts, which involves a bleaching process in the meaning, particularly with the inferential value of *eso* becoming more prominent.

In contrast to the previous examples, these cases cannot be segmented in the same manner according to the VAM. *Pues eso mismo te digo* or *Pues eso mismo* as parentheticals might possess a procedural meaning in dialogic contexts, contributing to the constructionalization process (see Cuenca 2017). However, these examples cannot be categorized as pure textual adjacent subacts (TASs), given that a conceptual basis is evident in the structure. *Pues* would be clearly tagged as a TAS, as summarized in the following schema comparing the first and second stages through the VAM:

- First stage of PE > # ACT # {DSS} (*pues eso* within the Directive Substantive Subact)
- Second stage of PE > Reactive intervention #ACT# {TAS} (*pues eso* as Textual Adjacent Subact)

To sum up, the agreement mark function of PE seems to be directly originated from *pues* used in clear dialogical contexts introducing responses since the 15th century (see

Iglesias 2000). These responses do not only agree with what is said but also allow for evaluating and adding further information to the previous intervention. This association of *eso* with *pues* in dialogic contexts will facilitate the evolution into more restrictive contexts linked to specialized responses geared towards explicit agreement. As observed in Figure 3, these uses of *pues* have been kept in the database until the present in coexistence with the constructionalized PE, triggered in the 19th century and consolidated in the 20th century (see also Section 4.2.3):

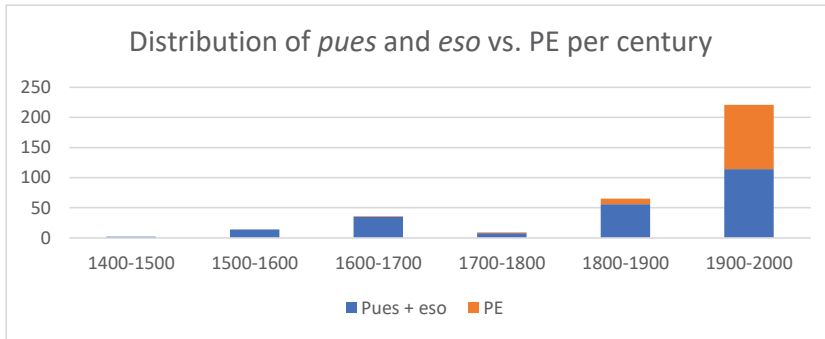


Figure 3. Distribution of *pues* and *eso* vs. PE per century in the database (Spain).

4.2.2.2. Development of the Reinforcement Value

The reinforcement value developed concurrently with the agreement mark value, as seen in Example (24):

(24) (...) derramado tu nombre’. Porque de todos aquestos nombres resulta un nombre, Jesús, de manera que no lo fuera ni se lo llamara, si alguno de ellos le faltara por caso. ¿Por ventura cada uno de nosotros no ve en sí y en la mudanza de sus voluntades que se llama Cristo Admirable? **Pues eso es ser Jesús.** Porque el principio de nuestra salud es, cuando comenzamos a aborrecer lo que antes amábamos, dolernos de lo que nos daba alegría, abrazarnos con lo que nos ponía temor, seguir lo que huíamos y desear con ansia lo que desechábamos con enfado.

... spilled your name.’ Because from all these names, one name emerges, Jesus, so that he would not be, nor would he be called, if any of them were missing. Do each of us not see in ourselves and in the change of our desires what is called Christ Admirable? Well, that is to be Jesus. Because the beginning of our salvation is when we start to detest what we once loved, feel pain for what brought us joy, embrace what filled us with fear, pursue what we used to flee from, and ardently desire what we previously rejected with annoyance.

(1583 LEÓN, FRAY LUIS DE, De los nombres de Cristo, libros I-III [España] [Eva Álvarez Martino, Oviedo, Universidad de Oviedo, 1996] Religión)

In (24), there is a monological context in which *pues* introduces a response to a preceding rhetorical question uttered by the same speaker. Morphosyntactically, this response, which includes *pues* and *eso*, forms a compact structure with a copula. There is limited syntactic flexibility and removing *pues* and *eso* would alter the propositional basis of the sentence. There is no semantic bleaching since the use of *pues* is connective, and *eso* works as a subject, referring to the immediate content introduced (*eso es ser Jesús*, En. *This is Jesus Christ*). In this context, the utterance can be interpreted literally, and no additional inferences are implied pragmatically. Nevertheless, a crucial point for analysis is the new position occupied by *pues* and not found in other previous monological texts, potentially

leading to further developments for this connective involving changes in scope and new functional uses combined with *eso*.

Example (25) exhibits similar features, with *eso* clearly referencing the preceding rhetorical question:

- (25) CENTENO Lo que ha de hacer es sufrir. ¿No sufrí yo lo del muerto? ¿Y burla de mí no hacía cuando después me corría? **Pues eso mismo** le advierto: ni se corra ni se asome hacia la parte que ofende, sufra como quien pretende y calle como quien come. CENTENO: What he must do is endure. Did I not endure what happened with the deceased? And did he not mock me when he later ridiculed me? Well, I give him the same advice: neither run nor lean towards the offending side, endure as one who intends to, and remain silent as one who eats.

(c1657 MORETO, AGUSTÍN, El Santo Cristo de Cabrilla [España] [Aurelio Valladares Reguero, Jaén, Real Sociedad Económica de Amigos del País, 2003] Verso)

Examples such as (25) appear later in our database (since the 17th century). They reflect a shift in *pues*, previously used in reproduced dialogues to introduce responses by different speakers: now, it mirrors a dialogic structure based on an adjacent “question–answer” pair apparently produced by different speakers but presented by the same writer/speaker. This shift also introduces a degree of polyphony and reflects an influence from the expletive value. In Example (25), *pues* and *eso* can accommodate additional elements that reinforce the anaphoric reference (e.g., the adjective *mismo*). As observed, the illative value is more relevant than the original, basic consequence meaning of *pues*, which is not activated in these new contexts. Unlike Example (24), modifying or even omitting the rest of the utterance in this context does not result in a complete loss of meaning.

A total of 35 examples spanning from the 16th to the 19th centuries display these characteristics, marking a preceding micro-stage in the development of reinforcement values. Specifically, the position occupied by *pues* in conjunction with *eso* creates room for the emergence of new pragmatic meanings related to intangible elements like ideas or concepts without explicit mention (see Section 4.2.3). This is corroborated by the analysis using the VAM, which clearly illustrates the modification affecting the form pole: apparently, *pues* occupies a medial position. A zoom through the VAM, however, shows an initial position of act; *pues* would be tagged, again, as a TAS.

This said, agreement and self-reinforcement functions in the second stage are experimenting a (semi)fixation, not a completed process. This is supported by the VAM since PE cannot be addressed as a free, adjacent unit (i.e., TAS, MAS, or IAS). The only item that can be tagged as a TAS is *pues*, showing specific modifications in the scope which are key to be fixed with *eso* as part of the same construction. There are no instances neither bridging formulation, back-to-topic, or closing mark uses in this stage.

4.2.3. Third Stage: Constructionalized PE

According to the data, the agreement and self-reinforcement values in progress in the second stage are established during the 19th century and consolidated in the 20th, especially during the second half of the century. This third stage is also related to post-constructionalization changes, which is supported by frequencies and by the development of new functions not previously detected, such as formulation and back-to-topic, as well as by the development of similar constructions based on *pues* (see Section 5).

Regarding the agreement mark, there are some first examples in the 17th century. However, the increase in frequencies takes place since ca. 1850. Observe Examples (26) and (27):

- (26) Franco: ¿Para qué es tan larga arenga?
 ¿Es más que hurtarle la dama
 y romperle la cabeza?
 Federico: Sí; que el ser público el caso

hace más viva la ofensa
y el descrédito mayor
que a darle muerte me empeña.
Franco: **Pues eso:** apretar la mano
y al sacudirle correrla.

Franco: Why is this speech so long? Is it more than just stealing the lady and breaking his head?

Federico: Yes, because making the matter public intensifies the offense and increases the discredit, which compels me to seek his death.

Franco: Well then, grip his hand tightly and, when shaking it, run it through.

(1652 MORETO, AGUSTÍN, *El lego del Carmen*. San Franco de Sena [España] [Florián Smieja, Salamanca, Ediciones Anaya, S.A., 1970] Verso)

(27) A: Eso es decencia. Murieron antes que vender el secreto del General. ¿Y dices que eran simples?

B: Como borregos.

A: Di que mártires, como los de Dios vivo.

B: **Pues eso.**

A: Los santos, ¿qué son?

A: That's decency. They died before selling the General's secret. And you say they were simple?

B: Like sheep

A: Say martyrs, like those of the living God.

B: **Pues eso.**

A: What are saints then?

(1898 PÉREZ GALDÓS, BENITO, *Zumalacárregui* [España] [Biblioteca Virtual Miguel de Cervantes, Alicante, Universidad de Alicante, 2002] Novela)

Example (26) was documented in the 17th century. This dialogical context shows a first parenthetical use of *pues eso* in which the meaning is not compositional and goes beyond the continuative value derived from *pues* in the second stage. As observed, Franco replies to Federico through *pues eso* and even introduces further content to the discourse, but this new content does not depend on *pues* and *eso*. Morphosyntactically, there is a wider reduction of the structure: from <*pues eso* + subordinate sentence with “que”> to <*pues eso* + verbal complements> and, finally, to <**PUES ESO**>. This agreement construction could still admit “to say” verbs (*pues eso digo*), but they are not very frequent in our database, especially in the 19th and 20th centuries. This structure does not admit negation (e.g., *pues no eso*), which also reveals a high degree of fixation.

Semantically, there is a bleaching on the original meaning of *pues* and *eso*, also derived from the acquisition of meanings from other elements they were combined with previously (*pues eso mismo*, *pues eso claro está*, verbs such as *decir*, etc.). Pragmatically, there are new invited inferences by which the interpretation is not literally an echoic repetition of what is said, but rather a clear intention of agreeing with the other speaker. In these contexts, *pues eso* could be substituted by agreement marks, such as *exacto*. Example (27) confirms this change: Speaker B replies to A through this PE construction with a new codified meaning: total agreement as answer in an adjacent pair. As mentioned earlier, the bleaching process appears to be complete, as the use of PE now allows reference to intangible elements, even those not previously mentioned.

As summarized in Table 5, the number of cases documented in our database has increased since the 20th century, suggesting a consolidation of these oral uses (see Table 4). The construction PE as an agreement mark still allows for some variation, including verbs (e.g., *pues eso digo*, *pues eso es lo que digo yo*, *pues eso es*), but this occurs only in a few cases in poems and novels, seemingly determined by the stylistic choices of their authors.

Table 5. PE as consolidated agreement mark in our database (Spain).

| | #Absolute | #Normalized (/1,000,000 Words) |
|--------------|-----------|--------------------------------|
| 15th Century | 0 | 0.00 |
| 16th Century | 0 | 0.00 |
| 17th Century | 1 | 0.03 |
| 18th Century | 0 | 0.00 |
| 19th Century | 8 | 0.23 |
| 20th Century | 105 | 0.97 |

The consolidation of the agreement function involves a new analysis with the VAM, reflecting changes in the form–meaning features through positions and units: contrary to first and second stages, PE as a parenthetical is tagged as interactive adjacent subact (IAS) in an independent intervention, with a new scope over the whole previous intervention. Also, it can be related to further content, thus tagged as an IAS placed at the initial position of the intervention. As observed, *pues* and *eso* are part of the same structure and cannot be addressed as different units (i.e., *pues* is an IAS and *eso* as part of a DSS, like in previous examples). The new position and unit acquired by PE are linked to its semantic bleaching and the new pragmatic readings developed, which also differ from previous cases employing *pues* as a continuative mark introducing responses.

Regarding the self-reinforcement function, the first examples of the completed construction are detected since 1900, as observed in (28):

- (28) en aquel rincón de la biblioteca donde están los pocos libros de la familia de tu madre (q. e. p. d.) busca una obra en cuatro tomos, en cuarto, de canto dorado, con el lomo muy pintado de arabescos, dorados también. Aquello es La leyenda de oro. **Pues eso.** ¿Te quejas? ¿Te parece ñoño, viejo, naif, el libro? ¿Qué dirías si te mandase buscar en el estante de los libros vetustos Leyenda áurea, por Santiago de Vorágine?

In that corner of the library where your mother’s family’s few books (may she rest in peace) are located, look for a four-volume work, in quarto size, with gilded edges and a spine adorned with intricate golden arabesques. That is ‘La leyenda de oro.’ Well, there you have it. Do you complain? Do you find the book too sentimental, old-fashioned, or naïve? What would you say if I asked you to find ‘Leyenda áurea’ by Santiago de Vorágine in the shelf of ancient books?

- (1901 CLARÍN (LEOPOLDO ALAS), Siglo pasado [España] [José Luis García Martín, Gijón, Llibros del Pexe, 1999])

In (28), PE shows features similar to example (27): morphosyntactically, it is a free item which can be replaced or deleted from the discourse without provoking a loss of meaning, and there is a clear semantic bleaching by PE expressing a meaning not coded only in the combination of *pues* and *eso* but also complemented by a new inferential meaning by which information or an opinion can be highlighted by the speaker himself. This function goes beyond the agreement mark or the answer value detected in the database during previous centuries. There is even a certain degree of liberation from the rhetorical question–answer structure found before, which constitutes the point of departure for this function, as observed in the next example:

- (29) Me lo decía ayer Miguel Juste, almorzando en Platerías: “USA, primera potencia material, se está convirtiendo en primera potencia espiritual”. **Pues eso.** Que Roca es potencia material en Cataluña y potencia espiritual en Madrid.

Miguel Juste told me this yesterday while having lunch at Platerías: “The USA, the leading material power, is becoming the leading spiritual power.” Well, there you have it. Roca is a material power in Catalonia and a spiritual power in Madrid.

(1 June 1984, El País, 1 June 1984 [España] [Madrid, Diario El País, S.A., 1 June 1984] Actualidad)

According to the VAM, PE would be placed at the final position of an act, over which it has scope, and tagged as a modal adjacent subact (MAS). Again, this involves a change in the parameters of the construction: PE working as an agreement mark can only be addressed as an IAS in a different position and with a different scope compared to self-reinforcement, oriented to the way the speaker put illocutionary force in the message.

Our database reveals the consolidation of PE in self-reinforcement contexts also during the 20th century (see Table 6). These data include the completely parenthetical stage of the construction, as a continuation of the data retrieved in the second stage (see Section 4.2.2.2). Interestingly, contexts of rhetorical question–answer continue being employed by contemporary speakers in Peninsular Spanish.

Table 6. PE as a self-reinforcement construction consolidated mark in our database (Spain).

| | #Absolute | #Normalized (1,000,000 Words) |
|--------------|-----------|-------------------------------|
| 15th Century | 0 | 0.00 |
| 16th Century | 0 | 0.00 |
| 17th Century | 0 | 0.00 |
| 18th Century | 0 | 0.00 |
| 19th Century | 1 | 0.02 |
| 20th Century | 42 | 0.38 |

Last, the formulation and the back-to-topic functions are detected as part of the complete constructionalization process since the 19th century onwards. The database, however, shows only a few cases of closing mark. Observe Examples (30) and (31):

(30) —¿Yo? Nada. Que somos novios. Y como es muy buena habrá querido decir. . . **pues eso**, que está contenta con lo nuestro.
—Me? Nothing. We’re dating. And since she’s very kind, she probably meant. . . well, that she’s happy with us.

(1966 MARSÉ, JUAN, Últimas tardes con Teresa [España] [Barcelona, Seix Barral, 1996] Novela)

(31) —Dña. Dolores: ¿Por qué dices eso? —María: A mi padre le fusilaron. Se enteró Basilio.
—Dña. Dolores: (Después de un silencio.) Y tú. . . ¿**piensas volver a trabajar?**
—María: **Sí, eso quería.** Pero me imagino que ustedes no estarán en situación. . . (Dirige una mirada interrogante a DON LUIS.) Nosotros, de momento. . .
—Don Luis: De momento, no.
—María: No, si ya me parecía a mí. Yo lo que había pensado es que como he sabido que se ha muerto la Josefa, la criada de doña María Luisa la casera, que iba a ir a pretender allí. Aunque yo no soy como la Josefa, porque yo no entiendo tanto de cocina, y es una casa más importante. (Rectifica, preocupada.) Bueno, sin hacer de menos.
—Dña. Dolores: No te preocupes, María; habla con tranquilidad.
—María: **Pues eso. . . que quería** preguntarles que qué les parece a ustedes que vaya a pretender a casa de doña María Luisa la casera.
—Dña. Dolores: Why do you say that?
—María: My father was executed. Basilio found out.
—Dña. Dolores: (After a pause.) And you. . . do you plan to go back to work?
—María: Yes, that’s what I wanted. But I imagine you all won’t be in a position. . . (She looks questioningly at DON LUIS.) At least, not for now. . .
—Don Luis: Not for now, no.
—María: Yeah, that’s what I thought. What I had in mind is that since I heard that Josefa, the maid of doña María Luisa the landlady, has passed away,

I was thinking of proposing there. Although I'm not like Josefa because I don't know as much about cooking, and it's a more important household. (She corrects herself, concerned.) Well, without diminishing her.

—Dña. Dolores: Don't worry, María; speak calmly.

—María: Well, that. . . I wanted to ask what you all think about me going to propose to doña María Luisa the landlady's house.

(1982 FERNÁN GÓMEZ, FERNANDO, *Las bicicletas son para el verano* [España] [Haro Tecglen, Eduardo, Madrid, Espasa Calpe, 1994] Teatro)

A total of 41 examples in the database, all extracted from novels and authentic discourses (interviews in radio or TV) from the 20th century onwards, exhibit the same structure as (30). These examples reflect a complete PE construction, including all the linguistic features associated with the third stage. This results in a new, reanalyzed form-meaning pairing oriented towards formulation (i.e., parenthetical, new meaning beyond the combination of *pues* and *eso*, employed in new pragmatic contexts, etc.). PE functions as a type of oral pause, allowing the speaker to think on the next idea to be introduced (e.g., *que está contenta con lo nuestro*). According to the VAM, formulation introduces new units and positions, specifically in the medial position within the uttered act, over which PE holds influence as a textual adjacent subact (TAS).

Beyond formulation, the most recent developments of PE identified in the database pertain to the back-to-topic function. As illustrated in Example (31), PE enables the speaker to return to a previously mentioned topic or idea that had been set aside following a digression. In this instance, Miss Dolores and Maria were discussing job opportunities at their home. After receiving a negative response and engaging in an extended discussion about Maria's qualifications as a housekeeper, Maria revisits the main topic regarding the possibility of working at Luisa's house. Although the number of back-to-topic examples in the database is limited, their documentation suggests that this function became established at a later stage. The VAM analysis also reveals functional distinctions compared to other functions of PE: it is once again categorized as a textual adjacent subact (TAS), but it is positioned at the initial position within the intervention, with influence over prior interventions.

This is similar to cases involving closing marks: there are no conversational instances in which PE is employed to conclude a first topic in the final position of a reactive intervention and dialogue (as discussed in Section 3). Instead, there are only five cases (from 1995 to 1997) that appear to exhibit this function, as the writer closes one topic and may open another:

(32) Y por eso nos llama tanto la atención. Toda novedad, que decía Nabucodonosor, según Borges, no es sino olvido. **Pues eso.** And that's why it catches our attention so much. Every novelty, as Nabucodonosor said, according to Borges, is nothing but forgetfulness. Well, there you have it.

(15 February 1996, *El Mundo*, 15 February 1996 [España] [Madrid, Unidad Editorial, 15 February 1996] Política)

In these cases, there is a reanalysis by which PE presents only textual properties organizing the discourse. According to the VAM, it would be tagged as a TAS in the final position of the dialogue and even discourse.

To sum up, the functional development of PE across the units and positions within the VAM enables us to create a visual representation of the constructionalization process completed in the third stage. Initially, the values of *pues* were confined to the act or the intervention, with a more limited scope (see Table 7):

Table 7. PE in the VAM grid.

| PUES ESO | Subact | Act | Intervention | Dialogue |
|-------------|--------|-------------------------|--|--------------------------|
| Initial | | | “pues” PE agreement (IAS) PE back-to-topic (TAS) | |
| Medial | “pues” | PE formulation (TAS) | | |
| Final | | PE self-reinf. (MAS) | | PE closing mark (TAS) |
| Independent | | | PE agreement (IAS) | |

Each unit and position, in turn, is associated with the set of form and meaning features described in Section 4.2. In turn, the evolutive pattern followed by PE in the third stage can be traced by including the date of the first documentation of each function developed in the combination unit × position (see Pons Bordería 2018 for further diachronic evolutions of discourse markers in Peninsular Spanish through the VAM). See Table 8:

Table 8. First documentations of PE in the VAM grid.

| PUES ESO | Subact | Act | Intervention | Dialogue |
|-------------|--------|------------------------|--|-------------------------|
| Initial | | | PE agreement 1885 PE back-to-topic 1982 | |
| Medial | | PE formulation 1959 | | |
| Final | | PE self-reinf. 1884 | | PE closing mark 1995 |
| Independent | | | PE agreement 1898 | |

As seen, PE abandons subacts and develops a new preference for medial and final positions in the 20th century, which differs from the original values of *pues* with a narrower scope and a more reduced positional distribution.

5. Discussion

The findings of the analysis show that PE becomes a complete conversational formula in the 20th century, but that there is a bridging change stage (i.e., the second stage) which is relevant for this constructionalization process. In particular, the agreement and self-reinforcement functions seem to be developed earlier before the consolidation of the par-enthetical combination of *pues* and *eso*. This second stage corresponds to pre-constructional changes leading to the new construction (e.g., contextual restrictions, changes in the morphosyntactic features, such as combinations with new complements and more specific verbs, new pragmatic readings, and gradual expansion, etc.). In addition, post-constructional changes take place in the third stage, as shown by the development of new functions derived from the new consolidated ones, not previously found in the database (i.e., formulation, back-to-topic, and closing mark), as well as by the clear loss of compositionality and structural reduction (Traugott and Trousdale 2013). Also, PE seems to be integrated into a bigger constructional network, as shown by other formulas such as *pues nada* or *pues vale*, commonly employed in colloquial, spoken discourses and also popularized during the 20th century (e.g., *pues nada*; Salameh Jiménez Forthcoming(a); see also *pues bien* in Octavio de Toledo 2018). The development of other similar constructions suggests a productive constructional schema behind *pues* which must be diachronically addressed.

As shown, constructionalization as a framework explains how conversational formulas such as PE are conventionalized (i.e., an increase in frequencies of use in the 19th and 20th

centuries in tandem with the acquisition of new form and meaning features), in line with other discursive devices which are not discourse markers but develop similar functions (e.g., *general extenders*; Borreguero Zuloaga 2023). While certain features of PE can be elucidated through grammaticalization, such as *reduction*, *decategorialization*, and *layering* in Hopper's framework, the complete spectrum of features associated with PE does not conform entirely to grammaticalization processes (e.g., there is no phonetic reduction; rather, there is a case of elision from complete clauses to truncated clauses; Brinton 2023, p. 25; in addition, the bleaching process is less transparent since the neutral deictic *eso* keeps some kind of referential meaning which cannot be completely erased). A constructional approach is also compatible with some basic principles of grammaticalization and would also explain the way *pues* and *eso* become PE: a "conversational formula" would be the more abstract level, composed of instances of the second stage of PE working as a (semi)fixed construction (e.g., *pues X* as an answer, rhetorical question–answer introduced by *pues*) which, in turn, includes fewer schematic formulas triggered later (e.g., *pues eso mismo le advierto*, *pues eso digo*, *pues eso es*, etc.), before the definitive rise of PE. These more abstract levels would also be related to the development of other *pues*-based constructions (Salameh Jiménez Forthcoming(a)).

Regarding the data, some specific aspects can also be discussed: this first diachronic analysis of PE allows for proposing specific periods explaining its evolution. These periods are based on the data retrieved from a representative corpus in Peninsular Spanish covering different centuries and examples from the 20th century (i.e., the CDH). Form and meaning changes detected in each stage can constitute the point of departure for deeper research on PE which requires new data sources. For example, some of the first examples of constructionalized PE belong to texts published by specific authors writing novels and plays reproducing oral interaction since the 19th century (e.g., Pérez Galdós, Aub, Sánchez Ferlosio, Pardo Bazán, etc.). Thus, a focus on this period is suggested so as to retrieve further examples of agreement and self-reinforcement functions in oral contexts and, especially, to check whether formulation, back-to-topic, or closing marks were triggered later or maybe before the 20th century. Similarly, examples from the second stage could also be explored from discourse traditions (Llopis-Cardona 2022) by addressing specific types of texts and stylistic features in writers employing *pues* and *eso* (e.g., religious texts; Traugott 2010). This, again, invites us to address a possible relationship between those first uses, probably idiosyncratic, and their diffusion.

Last, the analysis shows that the incorporation of the VAM units and positions as part of the form meaning allows for a clearer association with the semantic and pragmatic changes experimented by PE as a new construction (Pons Bordería 2018): the uses of *pues* and *eso* in the first stage are part of directive substantive subacts (DSSs) within acts; cases of *pues* and *eso* developing dialogical features in the second stage involve a new segmentation by which *pues* works as a textual adjacent subact (TAS) introducing the rest of the DSS in which *eso* is inserted; finally, PE parenthetical cases in the third stage can be tagged as interactive adjacent subacts (IASs) for the agreement function, modal adjacent subacts (MASs) for the self-reinforcement, and textual adjacent subacts (TASs) for formulation, back-to-topic, and closing marks. This suggests an interpersonal > modal > textual cline, in line with other works in different languages (Scivoletto 2023). These transformations in units are also combined with positional changes (from medial in subacts to initial in interventions, and medial and final in acts and even dialogues).

6. Conclusions

The diachronic study of PE in Peninsular Spanish has remained unexplored, mainly due to the spoken, colloquial nature of this conversational formula, which could hinder data obtention. However, a first analysis of the CDH (RAE) has shed some light on its evolution, with a total of 732 tokens classified into three stages according to the degree of fixation detected. By adopting a constructional approach (Traugott 2020), results reveal that the different current functions covered by PE were consolidated during the 20th century, with special attention to previous developments in the 16th and 17th centuries, when

agreement and self-reinforcement functions seem to have originated. In turn, formulation, back-to-topic, and closing marks were triggered later, as part of post-constructural changes.

Our results allow comparisons with the development of other similar discursive devices, such as *general extenders* (e.g., *y eso*, *y nada*, *y todo*; López Serena 2018) and fit results from other works dealing with grammaticalization processes finished in the 20th century (see Salameh Jiménez 2020b; Pardo-Llibrer Forthcoming). Some tasks, however, need future research to be done: first, the second and third stages require further data focusing on specific text types, especially in the 20th century with other complementary corpora to be explored (Biblioteca Virtual de Prensa Histórica; or spoken colloquial corpora, such as Val.Es.Co. in its initial era—i.e., 1990s—spontaneous interviews from PRESEEA, etc.). Additionally, the consolidation of formulation, back-to-topic, and closing marks could be checked through interviews with 60–80-year-old speakers, complementing the data retrieved. Second, a diachronic approach to other closer conversational formulas, such as *pues nada*, *pues bien*, or *pues vale*, is needed to complete the constructional network to which *pues* belongs, as well as to measure the schematicity and productivity behind PE as a possible influence for their development. Again, the use of the VAM units as part of the form pole will be useful in systematizing all the functional transformations experimented by (semi)completed constructions.

Funding: This research was funded by the projects CIPROM/2021/038 Hacia la caracterización diacrónica del siglo XX (DIA20), Generalitat Valenciana, and I+D+I PID2021-125222NB-IOO Aportaciones para una caracterización diacrónica del siglo XX, MCIN/AEI/10.13039/501100011033/ and FEDER, Una manera de hacer Europa.

Institutional Review Board Statement: The study did not require ethical approval. Ethical review and approval were waived for this study as it involved no experimenting with humans. The analysis was carried out through an already existing corpus from the Real Academia Española (Spain), which is publicly accessible online.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the examples retrieved for this analysis can be found at <https://apps.rae.es/CNDHE/>. Accessed on: 28 September 2023.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ The diachronic application of the VAM is possible, as shown by previous studies dealing with grammaticalization processes in discourse markers (Salameh Jiménez 2020a). Although the VAM was initially created to analyze colloquial conversations, further developments have been carried out to address other oral–written genres (interviews and social gatherings, Briz 2013; football broadcastings; Salameh Jiménez (Forthcoming(b)); and, more recently, written journalistic texts, Pons Bordería 2022).
- ² A search for “eso mismo” in the CDH to determine whether it functions as a consolidated formula in parallel with “pues eso mismo” reveals that there is no interdependence between them. “Eso mismo” is often combined with “E eso mismo” or “Y eso mismo”, but these combinations have a different, non-continuative meaning.

References

- Achard, Michel. 2001. French *ça* and the Dynamics of Reference. In *LACUS Forum 27: Speaking and Comprehending*. Edited by Ruth M. Brend, Alan K. Melby and Arle R. Lommel. Toronto: Linguistic Association of Canada and the United States, pp. 49–62.
- Álvarez, Alfredo. 1999. Las construcciones consecutivas. In *Gramática de la lengua española. Tomo III*. Directed by Ignacio Bosque and Violeta Demonte. Madrid: Espasa Calpe, p. 3741.
- Arguedas, María Estellés. 2011. *Gramaticalización y paradigmas. Un estudio a partir de los denominados marcadores de digresión en español*. Frankfurt am Main: Peter Lang.
- Boas, Hans. 2010. Comparing constructions across languages. In *Contrastive Studies in Construction Grammar*. Amsterdam and Philadelphia: John Benjamins, pp. 1–20.
- Borreguero Zuloaga, Margarita. 2023. La gramaticalización de los apéndices generalizadores en español. Fenómenos de diacronía del s. XX. *Boletín de Filología* 58: 211–41. [CrossRef]
- Brinton, Laurel. 2023. The rise of what-general extenders in English. *Journal of Historical Pragmatics*. [CrossRef]
- Briz, Antonio. 1993. Los conectores pragmáticos en la conversación coloquial (II): Su papel metadiscursivo. *Español Actual: Revista de español vivo* 59: 39–56.

- Briz, Antonio. 2013. Variación pragmática y coloquialización estratégica. El caso de algunos géneros televisivos (la tertulia). In *(Des)cortesía para el espectáculo: Estudios de pragmática variacionista*. Edited by Catalina Fuentes. Madrid: Arco Libros, pp. 89–125.
- Briz, Antonio, Antonio Hidalgo, Xose Padilla, Salvador Pons, Leonor Ruiz Gurillo, Julia Sanmarhn, Elisa Benavent, Marta Albelda, María José Eernández, and Montserrat Pérez. 2003. Un sistema de unidades para el estudio del lenguaje coloquial. *Oralia* 6: 7–61. [CrossRef]
- CDH. n.d. CDH from the Real Academia Española. Available online: <https://www.rae.es/banco-de-datos/cdh> (accessed on 28 September 2023).
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Croft, William, and Alan Cruse. 2008. *Lingüística Cognitiva*. Madrid: Editorial Akal.
- Cuenca, Maria Josep. 2017. Connectors gramaticals i connectors lèxics en la construcció discursiva del debat parlamentari. *Zeitschrift für Katalanistik* 30: 99–121. [CrossRef]
- De Cock, Barbara. 2013. Entre distancia, discurso e intersubjetividad: Los de mostrativos neutros en español. *Anuario de Letras: Lingüística y Filología* 1: 5–36. [CrossRef]
- Del Rey Quesada, Santiago. 2020. Hacia una diacronía de la oralidad: El inicio de turno y la inmediatez comunicativa en un corpus de traducciones de Plauto y Terencio (ss. XVI y XIX). *Lexis* 44: 41–74. [CrossRef]
- Deng, Delin. 2023. The Grammaticalization of the Discourse Marker genre in Swiss French. *Languages* 8: 28. [CrossRef]
- do Rosário, Ivo da Costa, and Mariângela Rios de Oliveira. 2016. Funcionalismo e abordagem construcional da gramática. *Alfa Revista de Lingüística* 60: 233–59. [CrossRef]
- Engiels, Renata, and Mar Garachana. 2021. Grammaticalization, lexicalization and constructionalization. In *The Routledge Handbook of Cognitive Linguistics*. Edited by Wen Xu and John R. Taylor. London: Routledge, pp. 314–32.
- Espinosa, Guadalupe. 2016. Dientes de sierra: Una herramienta para el estudio de la estructura interactiva del discurso dialógico. *Normas* 6: 13–27. [CrossRef]
- Estellés Arguedas, Maria, and Salvador Pons Bordería. 2014. Absolute Initial Position. In *Models of Discourse Segmentation in Romance Languages*. Edited by Salvador Pons Bordería. Amsterdam: John Benjamins, pp. 121–55.
- Fagard, Benjamin, and Alexandru Mardale. 2015. The pace of grammaticalization and the evolution of prepositional systems: Data from Romance. *Folia Linguistica* 46: 303–40. [CrossRef]
- Fillmore, Charles J. 1988. The mechanisms of ‘Construction grammar’. Paper presented at the Fourteenth Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, USA, February 14–15; vol. 14, pp. 35–55.
- Fischer, Kerstin, and Maria Alm. 2013. A radical construction grammar perspective on the modal particle-discourse particle distinction. In *Discourse Markers and Modal Particles: Categorization and Description*. Edited by Liesbeth Degand, Bert Cornillie and Paola Pietrandrea. Amsterdam and Philadelphia: John Benjamins, pp. 47–88.
- Flach, Susanne. 2018. Constructionalization and the Sorites Paradox. The emergence of the *into*-causative. In *Nodes and Networks in Diachronic Construction Grammar*. Edited by Lotte Sommerer and Elena Smirnova. Constructional Approaches to Language. Amsterdam: John Benjamins Publishing Company, vol. 27, pp. 45–68.
- Fried, Mirjam, and Jan-Ola Östman. 2004. Construction Grammar in a cross-language perspective. *Studies in Language* 31: 479–85.
- Garachana, Mar, and María Sol Sansiseña. 2023. Combinatorial Productivity of Spanish Verbal Periphrases as an Indicator of Their Degree of Grammaticalization. *Languages* 8: 187. [CrossRef]
- Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist’s field trip. *Chicago Linguistic Society* 7: 394–415.
- Gras, Pedro. 2011. *Gramática de Construcciones en Interacción. Propuesta de un modelo y aplicación al análisis de estructuras independientes con marcas de subordinación en español*. Barcelona: Universitat de Barcelona.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37: 1043–68. [CrossRef]
- Hopper, Paul. 1991. On some principles on grammaticalization. In *Approaches to Grammaticalization*. Edited by Elizabeth Closs Traugott and Bernd Heine. Amsterdam and Philadelphia: John Benjamins, vol. I, pp. 17–36.
- Hopper, Paul, and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge, MA: Cambridge University Press.
- Iglesias, Silvia. 2000. La evolución histórica de *pues* como marcador discursivo hasta el siglo XV. *Boletín de la Real Academia Española* LXXX: 209–307.
- Llopis-Cardona, Ana. 2020. Funciones, posición y unidades discursivas en no sé y yo qué sé. In *Aportaciones desde el español y el portugués a los marcadores discursivos treinta años después de Martín Zorraquino y Portolés*. Edited by Antonio Da Silva, Catalina Fuentes and Manuel Martí-Sánchez. Sevilla: Universidad de Sevilla, pp. 249–72.
- Llopis-Cardona, Ana. 2022. El cambio de entorno y evocación en las tradiciones discursivas: Propuesta y aplicación al caso de *en definitiva*. *ELUA: Estudios De Lingüística. Universidad De Alicante* 37: 185–220. [CrossRef]
- López Serena, Araceli. 2018. Hacia una revisión de la caracterización semántica y discursiva de la locución y eso que en español actual. *ELUA: Estudios De Lingüística. Universidad De Alicante* 32: 195–217. [CrossRef]
- López Serena, Araceli. 2023. Discourse traditions and variation linguistics. In *Manual of Discourse Traditions in Romance*, 1st ed. Edited by Esme Winter-Froemel and Álvaro S. Octavio de Toledo y Huerta. Berlin: De Gruyter Mouton, vol. 30, pp. 59–80.
- López Serena, Araceli, and Daniel Sáez Rivera. 2018. Procedimientos de mimesis de la oralidad en el teatro español del siglo XVIII. *Estudios Humanísticos: Filología* 40: 235–73. [CrossRef]
- Octavio de Toledo, Álvaro. 2018. Paradigmaticisation through Formal Resemblance: A History of the Intensifier bien in Spanish Discourse Markers. In *Beyond Grammaticalization and Discourse Markers*. Edited by Salvador Pons Bordería and Óscar Loureda Lamas. Studies in Pragmatics: New Issues in the Study of Language Change, 18. Leiden: Brill, pp. 160–97.

- Pardo-Llibrer, Adrià. Forthcoming. From vocative contexts to digressive uses. On Spanish 'ah'. In *From Intersubjective to Textual Meaning: Motivation for the Rise of Discourse Markers/Pragmatic Elements*. Edited by Giulio Scivoletto and Ryo Takamura. Oxford: Oxford University Press.
- Petré, Peter. 2020. How constructions are born. The role of patterns in the constructionalization of *be going to* INF. In *Patterns in Language and Linguistics*. Edited by Beatrix Busse and Ruth Möhlig-Falke. Berlin: De Gruyter Mouton, pp. 157–92.
- Pons Bordería, Salvador. 1998. *Conexión y conectores: Estudio de su relación en el registro formal de la lengua*. València: Publicacions de la Universitat de València.
- Pons Bordería, Salvador. 2016. Cómo dividir una conversación en actos y subactos. In *Oralidad y análisis del discurso: Homenaje a Luis Cortés Rodríguez*. Edited by Antonio Miguel Bañón Hernández, María del Mar Espejo Muriel, Bárbara Herrero Muñoz-Cobo and Juan Luis López Cruces. Almería: Universidad de Almería, pp. 545–66.
- Pons Bordería, Salvador. 2018. Introduction. New insights in Grammaticalization studies. In *Beyond Grammaticalization and Discourse Markers. New Issues in the Study of Language Change*. Edited by Salvador Pons Bordería and Óscar Loureda Lamas. Leiden: Brill, pp. 1–16.
- Pons Bordería, Salvador. 2022. *Creación y análisis de corpus orales: Saberes prácticos y reflexiones teóricas (Sprache-Gesellschaft-Geschichte)*. Frankfurt am Main: Peter Lang.
- Pons Bordería, Salvador. 2023. Corpus Val.Es.Co. 3.0. Available online: <http://www.valesco.es> (accessed on 28 September 2023).
- Pons Bordería, Salvador, and Kerstin Fischer. 2021. Using discourse segmentation to account for the polyfunctionality of discourse markers: The case of *well*. *Journal of Pragmatics* 173: 101–18. [CrossRef]
- Salameh Jiménez, Shima. 2020a. ¿Por qué el español peninsular no reformula con *digamos*? Una reflexión sincrónica y diacrónica a partir de su comparación con el portugués europeo. In *Marcadores discursivos. O português como referência contrastiva*. Edited by Isabel Duarte and Rogelio Ponce de Leon. Theorie und Vermittlung der Sprache. Frankfurt am Main: Peter Lang, vol. 63, pp. 312–41.
- Salameh Jiménez, Shima. 2020b. *Pues eso* como construcción interactiva desde el modelo Val.Es.Co. *Zeitschrift für Katalanistik* 33: 99–130. [CrossRef]
- Salameh Jiménez, Shima. Forthcoming(a). The consolidation of the 'pues'-based constructional network in Peninsular Spanish during the 20th century.
- Salameh Jiménez, Shima. Forthcoming(b). The study of colloquialization through the VAM units: A visual approach to football broadcastings.
- Scivoletto, Giulio. 2023. Il siciliano *bi e l'espressione della miratività*. *Cuadernos de Filología Italiana* 30: 189–206. [CrossRef]
- Sweetser, Eve E. 1988. Grammaticalization and semantic bleaching. Paper presented at the Fourteenth Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, USA, February 14–15; vol. 14, pp. 389–405. [CrossRef]
- Traugott, Elizabeth Closs. 1982. From Propositional to Textual and Expressive Meanings: Some Semantic-Pragmatic Aspects of Grammaticalization. In *Perspectives on Historical Linguistics*. Edited by Winfred P. Lehmann and Yakov Malkiel. Amsterdam and Philadelphia: Benjamins, pp. 245–71.
- Traugott, Elizabeth Closs. 1995. The role of the development of discourse markers in a theory of grammaticalization. Paper presented at the Twelfth International Conference on Historical Linguistics, Manchester, UK, August 13–18.
- Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (inter)subjectification: A Reassessment. In *Subjectification, Intersubjectification and Grammaticalization*. Edited by Kristin Davidse, Lieven Vandelanotte and Hubert Cuyckens. Berlin: De Gruyter Mouton, pp. 29–74.
- Traugott, Elizabeth Closs. 2015. Toward a coherent account of grammatical constructionalization. In *Diachronic Construction Grammar*. Edited by Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer and Spike Gildea. Amsterdam: John Benjamins, pp. 51–80.
- Traugott, Elizabeth Closs. 2020. Constructional pattern-development in language change. In *Patterns in Language and Linguistics*. Edited by Beatrix Busse and Ruth Möhlig-Falke. Berlin: De Gruyter Mouton, pp. 125–56.
- Traugott, Elizabeth Closs. 2022. *Discourse Structuring Markers in English. A Historical Constructionalist Perspective on Pragmatics*. Constructional Approaches to Language. Amsterdam: John Benjamins Publishing Company.
- Traugott, Elizabeth Closs, and Bernd Heine. 1991. *Approaches to Grammaticalization. Theoretical and Methodological Issues*. Amsterdam: John Benjamins Publishing Company, vol. 1.
- Traugott, Elizabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Traugott, Elizabeth Closs, and Richard B. Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Trousdale, Graeme. 2014. On the relationship between grammaticalization and constructionalization. *Folia Linguística* 48: 557–78. [CrossRef]
- Val.Es.Co. Group. 2002. *Corpus de Conversaciones Coloquiales*. Madrid: Arco Libros.
- Val.Es.Co. Group. 2014. Las unidades del discurso oral. La propuesta Val.Es.Co. de segmentación de la conversación coloquial. *Estudios de lingüística del español* 35: 13–73.
- Zhuo, Jing-Schmidt, and Xinjia Peng. 2016. The emergence of disjunction: A history of constructionalization in Chinese. *Cognitive Linguistics* 27: 101–36.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages



Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

[mdpi.com](https://www.mdpi.com)

ISBN 978-3-7258-2102-0